University of Alberta

Applications in Finance of Hidden Markov Models

by

Barbara Mary Jamieson

A thesis

submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Statistics

Department of Mathematical Sciences

Edmonton, Alberta

Fall 1996

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch
Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-18276-2

Canada

University of Alberta

Library Release Form

Name of Author: Barbara Mary Jamieson

Title of Thesis: Applications in Finance of Hidden Markov Models

Degree: Master of Science

Year this Degree Granted: 1996

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

B. M. Jamieson

(Signed)

Permanent Address:

1027 Strathcona Drive

Sherwood Park, Alberta

T8A 1A5

Date: September 27, 1996

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Applications in Finance of Hidden Markov Models** submitted by Barbara Mary Jamieson in partial fulfillment of the requirements for the degree of Master of Science in Statistics.

R.J. Elliott (Supervisor)

A. Cadenillas

R.J. Karunamuni

R.M. Korkie

Date: Sept. 18, 1996

To Mary, James Hilton, Margaret, and James Duncan

Abstract

This thesis uses hidden Markov models to analyze an asset pricing problem and to model the term structure of interest rates.

Chapter 1 reviews selected topics in mathematical finance. It discusses discrete-time and continuous-time models of financial markets and some models of interest rates.

Chapter 2 discusses hidden Markov models. In these models, a discrete-time, finite-state Markov chain is observed through a function whose values are distorted by noise.

In Chapter 3, a unit-delay model with a real-valued observation process is applied in three examples using IBM stock prices, gold prices, and United States-Canadian exchange rates.

In Chapter 4, a zero-delay model with a vector-valued observation process is applied in an analysis of the United States term structure of interest rates.

Two appendices give the computer programs written to implement the estimation procedures of Chapters 3 and 4.

ACKNOWLEDGEMENTS

I wish to thank my thesis supervisor, Dr. Elliott, for his kind and expert guidance through a new and exciting field. It is very much appreciated.

I would also like to thank Dr. Cadenillas, Dr. Karunamuni, and Dr. Korkie for reading my thesis and for their thoughtful comments on it.

Lastly, I express my gratitude to Dr. Al-Hussaini, Dr. Elliott, Dr. Gombay, Dr. Hooper, Dr. Prasad, Dr. Poliquin, and Dr. Wiens for exemplary classroom instruction and generous assistance on research projects.

Contents

1 Related Mathematical Finance Theory **1**

1.1 Discrete-Time Models 1

1.1.1 Price Dynamics 1

1.1.2 Trading Strategies 3

1.1.3 Viable and Complete Financial Markets 6

1.2 Continuous-Time Models 7

1.2.1 Price Dynamics 8

1.2.2 Trading Strategies 9

1.2.3 Viable and Complete Financial Markets 11

1.3 Models of Interest Rates 11

1.3.1 Bond Price Dynamics 12

1.3.2 The Cox-Ingersoll-Ross Model 15

2 Hidden Markov Models **18**

2.1 Preliminaries 18

2.2 Unit Delay Observation Model 21

2.2.1 Real-Valued Observation Process 21

2.2.2 Vector-Valued Observation Process 37

2.3 Zero Delay Observation Model 39

2.3.1 Real-Valued Observation Process 39
2.3.2 Vector-Valued Observation Process 42

3 Application: Asset Pricing **44**

3.1 Preliminaries 44
3.2 The Model 45
3.3 Price Prediction 46
3.4 Applications and Results 48
3.4.1 The Program 48
3.4.2 Starting Values 49
3.4.3 The Time Step 50
3.4.4 Checking the Model 51
3.4.5 Magnifying the Price Changes 51
3.4.6 Case 1: IBM Stock Prices 53
3.4.7 Case 2: Gold Prices 58
3.4.8 Case 3: United States Dollar Prices 62

4 Application: Modelling the Term Structure of Interest Rates **89**

4.1 Preliminaries 89
4.2 The Model 90
4.2.1 Instantaneous Rate Process 90
4.2.2 Bond Dynamics 91
4.2.3 Filtering and Model Estimation 95
4.3 Application and Results 97
4.3.1 The Program 97
4.3.2 The Size of N 97

4.3.3 Starting Values 98
4.3.4 The Time Step 98
4.3.5 Checking the Model 98
4.3.6 Scaling the Yield Values 99
4.3.7 Application of the Model 100
4.3.8 Concluding Remarks 103

References **119**

A Computer Program for Chapter 3 **120**

B Computer Program for Chapter 4 **121**

List of Tables

3.1 $\hat{g}_{15}$ and $\hat{\gamma}_{15}$, after the fifteenth pass, for selected values of N – IBM stock prices 67

3.2 Starting values and parameter estimates – IBM stock prices 68

3.3 The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – IBM stock prices 69

3.4 Regressions of actual prices on predicted prices – IBM stock prices 70

3.5 The term that is multiplied by S_{k-1} to get $E[S_k \mid \mathcal{Y}_{k-1}]$, calculated at the end of each pass through the data 71

3.6 $\hat{g}_5$ and $\hat{\gamma}_5$, after the fifteenth pass, for selected values of N – gold prices 72

3.7 Starting values and parameter estimates – gold prices 73

3.8 The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – gold prices 74

3.9 Regressions of actual prices on predicted prices – gold prices 75

3.10 $\hat{g}_5$ and $\hat{\gamma}_5$, after the fifteenth pass, for selected values of N – U.S. dollar prices 76

3.11 Starting values and parameter estimates – U.S. dollar prices . . . 77

3.12 The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – U.S. dollar prices 78

3.13 Regressions of actual prices on predicted prices – U.S. dollar prices 79

4.1 Starting values and parameter estimates – 1-, 3-, and 6-month Treasury bills and 2-year bond 105

4.2 Starting values and parameter estimates – 5-, 7-, 10-, and 30-year bonds 106

4.3 Regressions of actual yields on estimated yields – 1-, 3-, and 6-month Treasury bills and 2-year bond 107

4.4 Regressions of actual yields on estimated yields – 5-, 7-, 10-, and 30-year bonds 108

List of Figures

3.1 Actual and predicted prices of IBM stock 80

3.2 Mean of y versus g.hat – IBM stock prices 81

3.3 Variance of y versus gamma.hat – IBM stock prices 82

3.4 Actual and predicted prices of gold 83

3.5 Mean of y versus g.hat – gold prices 84

3.6 Variance of y versus gamma.hat – gold prices 85

3.7 Actual and predicted prices of U.S. dollar 86

3.8 Mean of y versus g.hat – U.S. dollar prices 87

3.9 Variance of y versus gamma.hat – U.S. dollar prices 88

4.1 Yields on 1-month T-bill 109

4.2 Yields on 3-month T-bill 110

4.3 Yields on 6-month T-bill 111

4.4 Yields on 2-year bond 112

4.5 Yields on 5-year bond 113

4.6 Yields on 7-year bond 114

4.7 Yields on 10-year bond 115

4.8 Yields on 30-year bond 116

Chapter 1

Related Mathematical Finance Theory

This chapter reviews selected topics in mathematical finance which are related to the concepts and models of Chapters 2, 3, and 4. It discusses discrete-time models of financial markets, continuous-time models of financial markets, and some models of interest rates. The motivation for much of the theory discussed in this chapter is to provide a foundation for the pricing of contingent claims.

1.1 Discrete-Time Models

This section discusses the discrete-time model of financial markets originally formulated by Harrison and Kreps [12] and Harrison and Pliska [13].

1.1.1 Price Dynamics

In the model, all processes are defined on a probability space $(\Omega, \mathcal{F}, P)$, where Ω, the sample space, has a finite number of elements each representing a possible state of the world, $\mathcal{F}$ is a σ-field in Ω, and P is a probability measure on $\mathcal{F}$.

Definition 1.1.1 *A* σ-field *defined on* Ω *is a class,* $\mathcal{F}$*, of subsets of* Ω *having the following properties:*

1. $\phi \in \mathcal{F}$;
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$*; that is,* $\mathcal{F}$ *is closed under complementation; and*
3. $A_1, A_2, \ldots \in \mathcal{F}$ *implies* $A_1 \bigcup A_2 \bigcup \ldots \in \mathcal{F}$*; that is,* $\mathcal{F}$ *is closed under the formation of countable unions.*

Definition 1.1.2 *A set function* P *on a* σ*-field* $\mathcal{F}$ *is called a* probability measure *if it satisfies the following conditions:*

1. $0 \leq P(A) \leq 1$ *for* $A \in \mathcal{F}$;
2. $P(\Omega) = 1$*; and*
3. *if* $A_1, A_2, \ldots$ *is a disjoint sequence of sets in* $\mathcal{F}$ *and if* $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$*, then* $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

Definition 1.1.3 *A* sub σ-field *is a collection of subsets of a* σ*-field, which is itself a* σ*-field.*

Definition 1.1.4 *On* $(\Omega, \mathcal{F}, \mathcal{P})$*, a* filtration *is an increasing family of sub* σ*-fields* $\mathcal{F}_n \subset \mathcal{F}$*. Increasing means that if* $n \leq m$*, then* $\mathcal{F}_n \subset \mathcal{F}_m$*;* $n, m \in \{0, 1, \ldots, N\}$.

The probability space $(\Omega, \mathcal{F}, P)$ is equipped with a filtration $\{\mathcal{F}_n\}$, where $n \in \{0, 1, \ldots, N\}$ and N is finite. Without loss of generality, we take $\mathcal{F}_0 = \{\phi, \Omega\}$ and $\mathcal{F}_N = \mathcal{F} = \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ denotes the power set of Ω. We also assume $P(\omega) > 0$ for all $\omega \in \Omega$.

The model considers a financial market with $D + 1$ securities.

Now, before proceeding further, we define some additional terms.

Definition 1.1.5 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *the function* $X : \Omega \to \mathbb{R}$, *is called a* random variable *if it is* $\mathcal{F}$*-measurable — that is, if* $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ *for every* $B \in \mathcal{B}(\mathbb{R})$, *where* $\mathcal{B}(\mathbb{R})$ *denotes the Borel* σ*-field in* $\mathbb{R}$.

Definition 1.1.6 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *a* stochastic process, $\{X_t; t \in T\}$, *is a collection of random variables indexed by* $t \in T$.

In what follows, T, the time set, equals $\{0, 1, \ldots, N\}$.

Note that the latter two definitions extend to random vectors.

On $(\Omega, \mathcal{F}, P)$, $S = \{S_n; n = 0, 1, \ldots, N\}$ is a $D+1$ dimensional stochastic process with $S_n = (S_n^0, S_n^1, \ldots, S_n^D) \in \mathbb{R}^{D+1}$, where S_n^d, $d = 0, 1, \ldots, D$, represents the price of the dth security at time n. Here, $S_n^d > 0$ and S_n^d is $\mathcal{F}_n$-measurable. The latter means that, at time n, investors know the present and past prices of the $D+1$ securities.

The security with superscript '0' represents riskless investments. Without loss of generality, suppose $S_0^0 = 1$. This defines a discount process $\beta = \{\beta_n\}$ by setting $\beta_n = \left(\frac{1}{S_n^0}\right)$. If the rate of interest on the riskless investments is positive, constant, and equal to r, then $S_n^0 = (1+r)^n$.

The model also defines the stochastic process $\tilde{S} = \{\tilde{S}_n; n = 0, 1, \ldots, N\}$ with $\tilde{S}_n = (\tilde{S}_n^0, \tilde{S}_n^1, \ldots, \tilde{S}_n^D) \in \mathbb{R}^{D+1}$, where $\tilde{S}_n^d = \beta_n S_n^d$, $d = 0, 1, \ldots, D$, represents the discounted price of the dth security at time n. Note $\tilde{S}_n^0 = 1$ for all n.

1.1.2 Trading Strategies

On $(\Omega, \mathcal{F}, P)$, $\phi = \{\phi_n; n = 0, 1, \ldots, N\}$ is a $D+1$ dimensional stochastic process with $\phi_n = (\phi_n^0, \phi_n^1, \ldots, \phi_n^D) \in \mathbb{R}^{D+1}$, where ϕ denotes an investor's

trading strategy and ϕ_n^d, $d = 0, 1, \ldots, D$, represents the quantity of the dth security held by the investor in his portfolio between times $n-1$ and n. The quantities ϕ_n^d may be positive or negative.

The process, ϕ, is assumed to be predictable.

Definition 1.1.7 *A trading strategy, ϕ, is said to be* predictable *if, for all $d \in \{0, 1, \ldots, D\}$, ϕ_0^d is $\mathcal{F}_0$-measurable and ϕ_n^d is $\mathcal{F}_{n-1}$-measurable, $n = 1, \ldots, N$.*

Thus if ϕ is predictable, it follows that the investor constructs his portfolio, ϕ_n, after the prices, S_{n-1}, are observed and holds this portfolio until after the prices, S_n, become known.

The value of the portfolio ϕ_n, just after it has been established at time $n-1$, is given by the inner product:

$$\langle \phi_n, S_{n-1} \rangle = \sum_{d=0}^{D} \phi_n^d S_{n-1}^d \ .$$

Its value at time n is:

$$V_n(\phi) = \langle \phi_n, S_n \rangle = \sum_{d=0}^{D} \phi_n^d S_n^d \ .$$

Thus, the change in value of the portfolio, due to changes in security prices between times $n-1$ and n, is:

$$\langle \phi_n, \Delta S_n \rangle = \sum_{d=0}^{D} \phi_n^d (S_n^d - S_{n-1}^d) \ .$$

Also, the present value of the portfolio at time n is:

$$\tilde{V}_n(\phi) = \beta_n \langle \phi_n, S_n \rangle = \langle \phi_n, \tilde{S}_n \rangle \ .$$

The model assumes, as well, that trading strategy ϕ is *self-financing*; that is, it requires that:

$$\langle \phi_n, S_n \rangle = \langle \phi_{n+1}, S_n \rangle \ , \tag{1.1}$$

$n = 1, \ldots, N-1$. This means that no funds are added to or withdrawn from the portfolio at any of the times $n = 1, \ldots, N$.

Remark 1.1.8 A trading strategy is self-financing if and only if any change in the value of the portfolio between times n and $n+1$ is entirely attributable to changes in security prices between times n and $n+1$ *or* to capital gains. From (1.1), we have:

$$V_{n+1}(\phi) - V_n(\phi) = \langle \phi_{n+1}, \Delta S_{n+1} \rangle \ , \ n = 1, \ldots, N-1 \ . \qquad (1.2)$$

□

Remark 1.1.9 It is straightforward to show that (1.2) is equivalent to each of the following:

$$1. \ \textit{for all } n \in \{1, \ldots, N\}, \ V_n(\phi) = V_0(\phi) + \sum_{i=1}^{n} \langle \phi_i, \Delta S_i \rangle \ , \qquad (1.3)$$

$$2. \ \textit{for all } n \in \{1, \ldots, N\}, \ \tilde{V}_n(\phi) = V_0(\phi) + \sum_{i=1}^{n} \langle \phi_i, \Delta \tilde{S}_i \rangle \ .$$

Thus, since $\sum_{i=1}^{n} \langle \phi_i, \Delta S_i \rangle$ represents the capital gains that the investor realizes on his portfolio up to time n, $n = 1, \ldots, N$, (1.3) indicates that a trading strategy is self-financing if and only if all changes in the value of the portfolio, between inception and time n, are due to capital gains on the investments.

□

The model also assumes that trading strategy ϕ is admissible.

Definition 1.1.10 *A trading strategy ϕ is said to be* admissible *if it is self-financing and $V_n(\phi) \geq 0$ for all $n \in \{0, 1, \ldots, N\}$.*

Let Φ denote the set of all admissible trading strategies.

1.1.3 Viable and Complete Financial Markets

Definition 1.1.11 *An* arbitrage opportunity *is some $\phi \in \Phi$ such that $V_0(\phi) = 0$ and $E[V_N(\phi)] > 0$, where E denotes expectation under P.*

Definition 1.1.12 *A market is said to be* viable *if no opportunities for arbitrage exist.*

As noted earlier, the motivation for much of the theory discussed here is to provide a foundation for the pricing of contingent claims.

Definition 1.1.13 *A* contingent claim *is a nonnegative, $\mathcal{F}_N$-measurable random variable X_N defined on the probability space $(\Omega, \mathcal{F}, P)$. It can be viewed as a contract or agreement that pays an amount $X_N(\omega)$ at time N if state ω prevails.*

Definition 1.1.14 *A contingent claim is said to be* attainable *if there exists a trading strategy, $\phi \in \Phi$, such that $V_N(\phi) = X_N$.*

Definition 1.1.15 *A market is said to be* complete *if every contingent claim is attainable.*

In what follows, we state the condition under which a market is viable and complete in the discrete-time model. First, however, we give some additional definitions.

Definition 1.1.16 *Suppose $\{\mathcal{F}_n\}$, $n \in \{0, 1, \ldots, N\}$, is a filtration of the measurable space $(\Omega, \mathcal{F})$ and $\mathcal{Z} = \{\mathcal{Z}_n\}$ is a stochastic process defined on $(\Omega, \mathcal{F})$ with values in $(E, \mathcal{E})$. Then $\mathcal{Z}$ is said to be* adapted *to $\{\mathcal{F}_n\}$ if $\mathcal{Z}_n$ is $\mathcal{F}_n$-measurable for each n.*

Definition 1.1.17 *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space with a filtration* $\{\mathcal{F}_n\}$, $n \in \{0, 1, \ldots, N\}$. *An adapted stochastic process* $M = \{M_n\}$ *is said to be a* martingale (*or* a martingale relative to the filtration $\{\mathcal{F}_n\}$) *if* $E[|M_n|] < \infty$ *and* $E[M_{n+m} \mid \mathcal{F}_n] = M_n$ *almost surely,* $m \in \{0, 1, \ldots, N - n\}$.

Note that these definitions extend to random vectors. We say that a sequence $M = \{M_n\}$ of random vectors in $\mathbb{R}^D$ is a martingale if each component of the vector M_n defines a real martingale.

Definition 1.1.18 *On the measurable space* $(\Omega, \mathcal{F})$, two probabilities P *and* Q are equivalent *if and only if, for all events* $A \in \Omega$, $P(A) = 0$ *if and only if* $Q(A) = 0$.

In the context of this model, Q equivalent to P signifies that $Q(\omega) > 0$ for all $\omega \in \Omega$.

Finally, the result which establishes the condition under which a viable market is complete follows.

Theorem 1.1.19 *A market is viable and complete if and only if there exists a unique probability measure* Q *equivalent to* P *under which the discounted prices of the securities are martingales.*

Proof: A proof of this result is given in Lamberton and Lapeyre [15], page 20. ∎

1.2 Continuous-Time Models

This section discusses the continuous-time model of financial markets originally formulated by Harrison and Pliska [13].

1.2.1 Price Dynamics

In the model, all processes are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ where $t \in [0,T]$ and $0 < T < \infty$. The *filtration*, $\{\mathcal{F}_t\}$, is an increasing family of sub σ-fields $\mathcal{F}_t \in \mathcal{F}$. *Increasing* means that if $s \leq t$, then $\mathcal{F}_s \subset \mathcal{F}_t$; $s,t \in [0,T]$. The following conditions are also assumed:

1. $\mathcal{F}_0 = \{A \subset \Omega \mid P(A) = 0\} \cup \Omega$;
2. $\{\mathcal{F}_t\}$ is *right continuous*, meaning that $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, $0 \leq t \leq T$; and
3. $\mathcal{F}_T = \mathcal{F}$.

The model considers a financial market with $D+1$ securities.

On $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, $S = \{S_t; 0 \leq t \leq T\}$ is a $D+1$ dimensional stochastic process with $S_t = (S_t^0, S_t^1, \ \ldots \ , S_t^D) \in \mathbb{R}^{D+1}$. The components of S, $S^0 = \{S_t^0\}$, $S^1 = \{S_t^1\}$, $\ldots$, $S^D = \{S_t^D\}$, are assumed adapted (meaning $S_t^d \in \mathcal{F}_t$ for $0 \leq t \leq T$), right continuous with left limits, and strictly positive. S_t^d represents the price of the dth security at time t.

The security with superscript '0' represents riskless investments. The price process, $S^0 = \{S_t^0\}$, of this security is modelled as:

$$S_t^0 = e^{\int_0^t r_s ds} \ , \ 0 \leq t \leq T \ ,$$

$$S_0^0 = 1 \ .$$

Here, $r = \{r_t\}$[1] represents the instantaneous rate process, an adapted random process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. Furthermore, a discount process $\beta = \{\beta_t\}$ is defined

[1] As Duffie and Kan[4] state:

> We may think of r_t as the interest rate at time t on loans of infinitesimal maturity. More properly, it is possible to invest one unit of account at any time t in deposits, and receive at any time $s \geq t$ the payoff $\exp\left(\int_t^s r_u du\right)$.

by setting $\beta_t = \left(\frac{1}{S_t^0}\right)$, $0 \le t \le T$.

The model also defines a discounted price process $\tilde{S} = \{\tilde{S}_t; 0 \le t \le T\}$, with components $\tilde{S}_t^1 = \{\tilde{S}_t^1\}, \ \ldots \ , \tilde{S}^D = \{\tilde{S}_t^D\}$. Here, $\tilde{S}_t^d = \beta_t S_t^d$ denotes the discounted price of the dth security at time t.

1.2.2 Trading Strategies

On $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, a trading stategy, $\phi = \{\phi_t; 0 \le t \le T\}$, is a $D+1$ dimensional stochastic process with $\phi_t = (\phi_t^0, \phi_t^1, \ \ldots \ , \phi_t^D) \in \mathbb{R}^{D+1}$. Here, ϕ_t^d represents the quantity of the dth security held by the investor in his portfolio at time t. The components of ϕ, $\phi^0 = \{\phi_t^0\}, \phi^1 = \{\phi_t^1\}, \ \ldots \ , \phi^D = \{\phi_t^D\}$, are assumed *locally bounded* and *predictable*. [2]

The value of the portfolio can again be expressed as a stochastic process $V(\phi) = \{V_t(\phi); 0 \le t \le T\}$, where:

$$V_t(\phi) = \langle \phi_t, S_t \rangle = \sum_{d=0}^{D} \phi_t^d S_t^d \ .$$

The discounted value process, $\tilde{V}(\phi) = \{\tilde{V}_t(\phi)\}$, is expressed as follows:

$$\tilde{V}_t(\phi) = \beta_t V_t(\phi) = \phi_t^0 + \sum_{d=1}^{D} \phi_t^d \tilde{S}_t^d \ .$$

The model assumes that trading strategy ϕ is *self-financing*; that is, it requires that:

$$\begin{aligned} V_t(\phi) &= V_0(\phi) + \int_0^t \phi_u dS_u \\ &= V_0(\phi) + \sum_{d=0}^{D} \int_0^t \phi_u^d dS_u^d \ , \ 0 \le t \le T \ . \end{aligned}$$

For a discussion of stochastic integrals such as $\int_0^t \phi dS$, see for example Elliott[5].

[2] Definitions of these terms, for the continuous-time model, are provided in Harrison and Pliska [13], pages 233-234.

Remark 1.2.1 It can be demonstrated that ϕ is self-financing if and only if the following condition holds: [3]

$$\begin{aligned} \tilde{V}_t(\phi) &= \tilde{V}_0(\phi) + \int_0^t \phi_u d\tilde{S}_u \\ &= \tilde{V}_0(\phi) + \sum_{d=1}^{D} \int_0^t \phi_u^d d\tilde{S}_u^d \ , \ 0 \leq t \leq T \ . \end{aligned} \tag{1.4}$$

□

The model also assumes that the trading strategy ϕ is admissible.

Definition 1.2.2 *A trading strategy ϕ is said to be* admissible *if $\tilde{V}(\phi) \geq 0$, condition* (1.4) *holds, and $\tilde{V}(\phi)$ is a martingale under Q, where Q is a probability measure equivalent to P.* [4]

The definitions of the terms *adapted process* and *martingale*, for the continuous-time case, are similar to those given earlier.

Definition 1.2.3 *Suppose $\{\mathcal{F}_t\}$, $t \in [0,T]$, is a filtration of the measurable space $(\Omega, \mathcal{F})$ and $\mathcal{Z} = \{\mathcal{Z}_t\}$ is a stochastic process defined on $(\Omega, \mathcal{F})$ with values in $(E, \mathcal{E})$. Then $\mathcal{Z}$ is said to be* adapted *to $\{\mathcal{F}_t\}$ if $\mathcal{Z}_t$ is $\mathcal{F}_t$-measurable for each t.*

Definition 1.2.4 *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space with filtration $\{\mathcal{F}_t\}$, $t \in [0,T]$. An adapted stochastic process $M = \{M_t\}$ is said to be a* martingale *(or a* martingale relative to the filtration $\{\mathcal{F}_t\}$ *) if $E[|M_t|] < \infty$ for all t and $E[M_t \mid \mathcal{F}_s] = M_s$, $s \leq t$.*

Let Φ denote the set of all admissible trading strategies.

[3] See Harrison and Pliska [13], Proposition 3.24, page 238.

[4] This definition is given in Harrison and Pliska [13], pages 239-240.

1.2.3 Viable and Complete Financial Markets

As in the discrete-time model, we are interested in establishing the condition under which a financial market is viable and complete. Our definitions of *viable market* and *completeness* are identical to those given earlier. However, the definitions of *arbitrage opportunity*, *contingent claim*, and *attainable contingent claim* require some minor modifications for the continuous-time case.

Definition 1.2.5 *An* arbitrage opportunity *is some $\phi \in \Phi$ such that $V_0(\phi) = 0$ and $E[V_T(\phi)] > 0$, where E denotes expectation under P.*

Definition 1.2.6 *A* contingent claim *is a positive, $\mathcal{F}_T$-measurable random variable X_T defined on the probability space $(\Omega, \mathcal{F}, P)$.*

Definition 1.2.7 *A contingent claim is said to be* attainable *if there exists a trading strategy, $\phi \in \Phi$, such that $V_T(\phi) = X_T$.*

Finally, we give a result for the continuous-time model analogous to Theorem 1.1.19 .

Theorem 1.2.8 *A market is viable and complete if and only if there exists a unique probability measure Q equivalent to P under which the discounted prices of the securities are martingales.*

Proof: See Harrison and Pliska [13], Corollary 3.36, page 241. ∎

1.3 Models of Interest Rates

Models of interest rates are used primarily for the pricing and hedging of bonds and options on bonds. In what follows, we discuss these models briefly, with specific reference to the model of Cox, Ingersoll, and Ross [3].

1.3.1 Bond Price Dynamics

We model the instantaneous rate process $r = \{r_t\}$, $t \in [0,T]$, $0 < T < \infty$, as an adapted random process on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. Here, $\{\mathcal{F}_t\}$ represents the complete filtration generated by Brownian motion.

Definition 1.3.1 *A* Brownian motion *is a stochastic process* $\{w_t; t \geq 0\}$, *on some probability space* $(\Omega, \mathcal{F}, P)$, *with the following properties:*

1. *the process starts at zero:*

$$w_0 = 0 \ a.s. \ ;$$

2. *its sample paths are continuous:*

   the map $t \to w_t(\omega)$ *is continuous for almost all* $\omega \in \Omega$; *and*

3. *it has stationary normally distributed, independent increments:*

   for $0 \leq s < t$, *the increment* $w_t - w_s$ *is normally distributed with mean 0 and variance* $t - s$ *and is independent of* $\mathcal{F}_s = \sigma\{w_u : u \leq s\}$.

Definition 1.3.2 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *the* σ-field generated by the random variables X_s, $s \leq t$, *where* $s, t \in [0,T]$, *is the family of sets* $X_s^{-1}(B) = \{\omega \in \Omega : X_s(\omega) \in B\}$, $B \in \mathcal{B}(\mathbb{R})$.

In this context, if $\mathcal{F}_t^0$ denotes the σ-field generated by the random variables w_s , $s \leq t$, then $\{\mathcal{F}_t\}$ *complete* means that $A \subset B$, $B \in \mathcal{F}_t^0$, and $P(B) = 0$ together imply that $A \in \mathcal{F}_t$. $\{\mathcal{F}_t\}$ also has the three properties listed earlier in section 1.2.1 .

In this model, we suppose that two classes of assets, a savings account and zero-coupon bonds, are traded. The price of the savings account is modelled as an adapted process $S = \{S_t^0\}$ where:

$$S_t^0 = e^{\int_0^t r_s ds} .$$

$$S_0^0 = 1 .$$

It is assumed that $\int_0^T |r_t| dt < \infty$ almost surely. The prices of the bonds are modelled as an adapted process $P = \{P(t,u);\ 0 \le t \le u\}$ where $P(t,u)$ denotes the price at time t of a zero-coupon bond with expiry date $u \le T$.

If the market is viable, it follows from the earlier results for continuous-time models, that there exists a probability measure Q equivalent to P under which, for all real $u \in [0,T]$, the discounted price process $\tilde{P} = \{\tilde{P}(t,u);\ 0 \le t \le u\}$ is a martingale. Here, $\tilde{P}(t,u) = \exp(-\int_0^t r_s ds)\ P(t,u)$.

Assuming viability and using $P(u,u) = 1$, we get the following result:

$$P(t,u) = E_Q \left(e^{-\int_t^u r_s ds} \mid \mathcal{F}_t \right) \tag{1.5}$$

where E_Q denotes expectation under Q.

Before proceeding further, we give a number of definitions and results.

Definition 1.3.3 *On $(\Omega, \mathcal{F})$, probability measure Q is* absolutely continuous *with respect to probability measure P if for each set A in $\mathcal{F}$, $P(A) = 0$ implies $Q(A) = 0$.*

Theorem 1.3.4 *On $(\Omega, \mathcal{F})$, Q is absolutely continuous with respect to P if and only if there exists a nonnegative random variable, ξ , such that:*

$$\textit{for all } A \in \mathcal{F},\ \ Q(A) = \int_A \xi\ dP .$$

Proof: See Lamberton and Lapeyre [15], page 73. ∎

Remark 1.3.5 We can observe the following:

1. The implication from left to right in Theorem (1.3.4) is the Radon-Nikodym theorem. [5]

2. ξ is called both the Radon-Nikodym derivative and the density of Q with respect to P. It is sometimes denoted $\frac{dQ}{dP}$.

□

Theorem 1.3.6 (Girsanov) *Consider the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, $t \in [0,T]$, where $\{\mathcal{F}_t\}$ is the complete filtration generated by Brownian motion. Let $\{f_t\}$ be an adapted process such that*

$$\int_0^T f_t^2 dt < \infty \ a.s.$$

and such that the process $\{\xi_t(f)\}$ defined by:

$$\xi_t(f) = \exp\left(\int_0^t f_s dw_s - \frac{1}{2}\int_0^t f_s^2 ds\right)$$

is a martingale. Then under the probability $P^{(\xi)}$ defined by $\frac{dP^{(\xi)}}{dP} = \xi_T(f)$, the process $\{b_t\}$, where $b_t = w_t + \int_0^t f_s ds$, is a Brownian motion.

Proof: See Elliott [5]. ∎

Now we let ξ_T denote the density of Q with respect to P. Because of the assumption that we have made on the filtration $\{\mathcal{F}_t\}$, we get the following results.

[5] A statement and proof of the Radon-Nikodym theorem are given in Billingsley [2], pages 443-445.

Lemma 1.3.7 *There exists an adapted process* $\{f_t\}$, $0 \leq t \leq T$, *such that for all* $t \in [0,T]$,

$$\xi_T = \exp\left(\int_0^t f_s dw_s - \frac{1}{2}\int_0^t f_s^2 \, ds\right)$$

almost surely.

Proof: A proof is given in Lamberton and Lapeyre [15], page 115 and Artzner and Dalbaen [1], page 106. ∎

Corollary 1.3.8

$$P(t,u) = E\left(e^{-\int_t^u r_s ds + \int_t^u f_s dw_s - \frac{1}{2}\int_t^u f_s^2 ds} \mid \mathcal{F}_t\right) \tag{1.6}$$

where E *denotes expectation under* P.

Proof: A proof is given in Lamberton and Lapeyre [15], page 115 and Artzner and Dalbaen [1], page 107. ∎

Equations (1.5) and (1.6) show that to calculate the price of bonds, we need to know either the evolution of r_t under Q or the evolution of the pair (r_t, f_t) under P. The model of Cox, Ingersoll, and Ross, which we examine next, describes the evolution of r_t under P by an equation of diffusion and then chooses the form of f_t so as to preserve the same type of equation under Q.

1.3.2 The Cox-Ingersoll-Ross Model

In this model, which is discussed in [3], the evolution of the instantaneous interest rate, r_t, is expressed as follows:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}\, dw_t$$

where κ, θ, and σ are positive, real-valued constants and $w = \{w_t\}$ represents a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$. As a result of the specification of the drift term, $\kappa(\theta - r_t)$, the process $\{r_t\}$ is mean-reverting; that is, the interest rate r_t is pulled toward a long-term value, θ, at a rate which is governed by the parameter κ. Also, the specification of the diffusion coefficient, $\sigma\sqrt{r_t}$, precludes negative interest rates.

Applying equation (1.5), it can be demonstrated that $P(r_t, t, u)$, the price at time t, $0 \leq t \leq T$, of a zero-coupon bond with expiry date $u \leq T$, is [6]:

$$\begin{aligned} P(r_t, t, u) &= E_Q\left(e^{-\int_t^u r_s ds} \mid \mathcal{F}_t\right) \\ &= A(t,u)\, e^{-B(t,u)\, r_t} \end{aligned}$$

where:

$$A(t,u) \equiv \left[\frac{2\gamma e^{[(\kappa+\lambda+\gamma)(u-t)]/2}}{(\gamma+\kappa+\lambda)(e^{\gamma(u-t)}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2},$$

$$B(t,u) \equiv \frac{2(e^{\gamma(u-t)}-1)}{(\gamma+\kappa+\lambda)(e^{\gamma(u-t)}-1)+2\gamma}, \text{ and}$$

$$\gamma \equiv ((\kappa+\lambda)^2 + 2\sigma^2)^{\frac{1}{2}} .$$

The term structure of interest rates at time t is given by [7]:

$$R(r_t, t, u) = [r_t B(t,u) - \log A(t,u)]/(u-t) .$$

Furthermore, of interest to the application reported in Chapter 4 of this thesis, are the results of Pearson and Sun [18]. They test a version of the two-factor Cox-Ingersoll-Ross (CIR) model, which is an extension of the one-factor model discussed above. [8] Based on the method of maximum likelihood,

[6] See Cox, Ingersoll, and Ross [3], page 393.

[7] See Cox, Ingersoll, and Ross[3], page 394.

[8] The two-factor CIR model is discussed in Section 7 of [3]. Whereas in the one-factor model, bond prices and the term structure of interest rates are determined by the instantaneous interest rate, in the two-factor model a second state variable, the price level, also drives the evolution of these variables.

they estimate the parameters of the CIR model using data on discount and coupon bonds. They reject the CIR model by the likelihood ratio test. They also assess the performance of the model in predicting the yield rates on bonds. They conclude that it performs as poorly as a naive "martingale model", which assumes that expected future yields are the current yields. This is of interest to the analysis of both Chapters 3 and 4.

Chapter 2

Hidden Markov Models

This chapter discusses hidden Markov models which will, in Chapters 3 and 4, be applied in the analysis of two financial problems. It considers a discrete-time, finite-state Markov chain which is observed through a function whose values are distorted by noise. [1] The chapter considers both real-valued and vector-valued observation processes and examines models that incorporate both a unit delay and a zero delay. It begins with a discussion of concepts and processes on which the model is based.

2.1 Preliminaries

All processes are defined on a complete probability space, $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω denotes a set of points, $\mathcal{F}$ a σ-field in Ω, and P a probability measure on $\mathcal{F}$. *Completeness* means that $A \subset B$, $B \in \mathcal{F}$, and $P(B) = 0$ together imply that $A \in \mathcal{F}$. Some additional definitions follow.

[1] Important references on this subject are Elliott, Aggoun, and Moore [8], Elliott [7], and Elliott [6].

Definition 2.1.1 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *the function* $X : \Omega \to \mathbb{R}^N$, $N \in \mathbb{N}$, *is called a* random variable *if it is* $\mathcal{F}$*-measurable — that is, if* $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ *for every* $B \in \mathcal{B}(\mathbb{R}^N)$, *where* $\mathcal{B}(\mathbb{R}^N)$ *denotes the Borel* σ*-field in* $\mathbb{R}^N$.

Definition 2.1.2 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *a* stochastic process, $\{X_t; t \in T\}$, *is a collection of random variables indexed by* $t \in T$.

In what follows, T, the time set, represents discrete time.

On the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, we suppose that the stochastic process $X = \{X_k; k \in \mathbb{N}\}$ is a Markov chain taking on values in a finite set $S = \{s_1, \ldots, s_N\}$. We consider the function ϕ_i, $1 \leq i \leq N$, defined by $\phi_i(s_j) = \delta_{ij}$ and the vector function $\phi(s) := (\phi_1(s), \ldots, \phi_N(s))'$, where $'$ denotes transpose. Then ϕ is a bijection from S onto the set $\Sigma = \{e_1, \ldots, e_N\}$ of unit vectors of $\mathbb{R}^N$. Using such a bijection, we can, without any loss of generality, take the state space of the Markov chain to be the set Σ.

We assume that X_0 is given, or that its distribution is known. We also assume that X is a time homogeneous Markov chain, so

$$P(X_{k+1} = e_j \mid \mathcal{F}_k) = P(X_{k+1} = e_j \mid X_k), \; k = 0, 1, \ldots,$$

where $\mathcal{F}_k^0 = \sigma\{X_0, \ldots, X_k\}$ denotes the σ-field generated by $X_0, \ldots, X_k$ and $\{\mathcal{F}_k\}$ represents the complete filtration generated by $\mathcal{F}_k^0$. Again, completeness means that $A \subset B$, $B \in \mathcal{F}_k^0$, and $P(B) = 0$ together imply that $A \in \mathcal{F}_k$.

Writing $a_{ji} = P(X_{k+1} = e_j \mid X_k = e_i)$, we let $A = (a_{ji})$, $1 \leq i, j \leq N$, $\sum_{j=1}^N a_{ji} = 1$, represent the transition matrix of X.

Definition 2.1.3 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *the* σ-field generated by the random variables

$X_0, \ldots, X_k$ *is the family of sets* $X_i^{-1}(B) = \{\omega \in \Omega : X_i(\omega) \in B\}$, $B \in \mathcal{B}(\mathbb{R}^N)$, $i = 0, 1, \ldots, k$.

Definition 2.1.4 *On* $(\Omega, \mathcal{F}, \mathcal{P})$, *a* filtration *is an increasing family of sub* σ*-fields* $\mathcal{F}_k \subset \mathcal{F}$. *Increasing means that if* $k \leq l$, *then* $\mathcal{F}_k \subset \mathcal{F}_l$; $l, k \in \mathbb{N}$.

Definition 2.1.5 *A* sub σ-field *is a collection of subsets of a* σ*-field, which is itself a* σ*-field.*

Definition 2.1.6 *Suppose* $\{\mathcal{F}_k\}$, $k \in \mathbb{N}$, *is a filtration of the measurable space* $(\Omega, \mathcal{F})$ *and* $Z = \{\mathcal{Z}_k\}$ *is a stochastic process defined on* $(\Omega, \mathcal{F})$ *with values in* $(E, \mathcal{E})$. *Then* Z *is said to be* adapted *to* $\{\mathcal{F}_k\}$ *if* $\mathcal{Z}_k$ *is* $\mathcal{F}_k$*-measurable for each* k.

Definition 2.1.7 *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space with a filtration* $\{\mathcal{F}_k\}$, $k \in \mathbb{N}$. *An adapted stochastic process* $M = \{M_k\}$ *is said to be a* martingale (*or* a martingale relative to the filtration $\{\mathcal{F}_k\}$) *if* $E[|M_k|] < \infty$ *and* $E[M_{k+l} \mid \mathcal{F}_k] = M_k$ *almost surely,* $l \in \mathbb{N}$.

Lemma 2.1.8 *The Markov chain* X *has the representation:*

$$X_{k+1} = AX_k + M_{k+1},$$

where $M = \{M_k\}$ *is a sequence of vector-valued* $\{\mathcal{F}_k\}$ *martingale increments.*

Proof: The result will follow if we can show that $E[M_{k+1} \mid \mathcal{F}_k] = \underline{0} \in \mathbb{R}^N$.

Now:

$$\begin{aligned} E[M_{k+1} \mid \mathcal{F}_k] &= E[X_{k+1} - AX_k \mid \mathcal{F}_k] \\ &= E[X_{k+1} \mid \mathcal{F}_k] - E[AX_k \mid \mathcal{F}_k] \\ &= E[AX_k \mid X_k] - AX_k \\ &= \underline{0}\,. \end{aligned}$$

■

We suppose that the X process is not observed directly, but rather there is an observation process $y = \{y_k; k \in \mathcal{Z}^+\}$ that provides noisy observations of X.

2.2 Unit Delay Observation Model

2.2.1 Real-Valued Observation Process

In this section, we consider the observation model:

$$X_{k+1} = AX_k + M_{k+1}$$
$$y_{k+1} = g(X_k) + \gamma(X_k)b_{k+1},\ y_{k+1} \in \mathbb{R},\ k \in \mathbb{N}.$$

Note that there is a unit delay between the state X at time k and the observation y at time $k+1$. Here, $\{b_k\}$ is a sequence of independent and identically distributed (i.i.d.) $N(0,1)$ random variables. Because $X_k \in \Sigma$, the functions g and γ are determined by vectors $g = (g_1, \ \ldots \ , g_N)'$ and $\gamma = (\gamma_1, \ \ldots \ , \gamma_N)'$ in $\mathbb{R}^N$; that is, $g(X_k) = \langle g, X_k \rangle$ and $\gamma(X_k) = \langle \gamma, X_k \rangle$ where $\langle , \rangle$ denotes the inner product in $\mathbb{R}^N$.

We assume that $\gamma_i \neq 0$ and thus, without any loss of generality, that $\gamma_i > 0$, $1 \leq i \leq N$. Also, on $(\Omega, \mathcal{F}, \mathcal{P})$, we let $\mathcal{F}_k^0 = \sigma\{X_0, \ \ldots \ , X_k\}$, $\mathcal{Y}_k^0 = \sigma\{y_1, \ \ldots \ , y_k\}$, and $\mathcal{G}_k^0 = \sigma\{X_0, \ \ldots \ , X_k, \ y_1, \ \ldots \ , y_k\}$. Then $\mathcal{F}_k$, $\mathcal{Y}_k$, and $\mathcal{G}_k$ represent, respectively, the completions of $\mathcal{F}_k^0$, $\mathcal{Y}_k^0$, and $\mathcal{G}_k^0$; all are contained in $\mathcal{F}$.

Change of Measure:

Suppose $b(.)$ is a real random variable with density $\phi(b)$ and g and γ are known constants. Let $y(.) = g + \gamma b(.)$.

We wish to introduce a new probability measure $\overline{P}$, using a density λ, so that $\left(\frac{d\overline{P}}{dP}\right) = \lambda$, and under $\overline{P}$ the random variable y has density ϕ.[2] That is,

$$\begin{aligned}
\overline{P}(y \leq t) &= \int_{-\infty}^{t} \phi(y)dy \\
&= \int_{\Omega} I_{y \leq t} d\overline{P} \\
&= \int_{\Omega} I_{y \leq t} \lambda dP \\
&= \int_{-\infty}^{\infty} I_{b \leq \frac{t-g}{\gamma}} \lambda(b)\phi(b)db \\
&= \int_{-\infty}^{t} \lambda(b)\phi(b)\frac{dy}{\gamma} .
\end{aligned}$$

It follows that $\lambda(b) = \frac{\gamma\phi(y)}{\phi(b)}$.

In our observation model, the b_k are i.i.d. $N(0,1)$ random variables. We let $\phi(.)$ represent the $N(0,1)$ density,

$$\begin{aligned}
\lambda_n &= \frac{\langle \gamma, X_{n-1} \rangle \phi(y_n)}{\phi(b_n)}, \; n \in \mathbb{N}, \\
\Lambda_0 &= 1, \; \textit{and} \\
\Lambda_k &= \prod_{n=1}^{k} \lambda_n, \; n \geq 1.
\end{aligned}$$

We then define a new probability measure $\overline{P}_k$ on $(\Omega, \mathcal{G}_k)$ by setting the restriction of the Radon-Nikodym derivative to $\mathcal{G}_k$ equal to Λ_k; that is, $\left(\frac{d\overline{P}_k}{dP} \mid \mathcal{G}_k\right) = \Lambda_k$.

Lemma 2.2.1 *$\{\Lambda_k\}$ is a $\{\mathcal{G}_k\}$ martingale under P.*

Proof: We must show that $E[\Lambda_{k+1} \mid \mathcal{G}_k] = \Lambda_k$.

[2]This analysis is based on Elliott, Aggoun, and Moore [8], page 60.

Now $E[\Lambda_{k+1} \mid \mathcal{G}_k]$

$$\begin{aligned}
&= E\left[\Lambda_k \frac{\gamma(X_k)\phi(y_{k+1})}{\phi(b_{k+1})} \mid \mathcal{G}_k\right] \\
&= \Lambda_k \gamma(X_k) E\left[\frac{\phi(y_{k+1})}{\phi(b_{k+1})} \mid \mathcal{G}_k\right] \\
&= \Lambda_k \gamma(X_k) E\left[\frac{\phi(g(X_k)+\gamma(X_k)b_{k+1})}{\phi(b_{k+1})} \mid \mathcal{G}_k\right] \\
&= \Lambda_k \gamma(X_k) \int_{\mathbf{R}} \frac{\phi(g(X_k)+\gamma(X_k)b)}{\phi(b)} \phi(b) dP \\
&= \Lambda_k \gamma(X_k) \int_{\mathbf{R}} \phi(z) \frac{dz}{\gamma(X_k)} \quad (z = g(X_k)+\gamma(X_k)b) \\
&= \Lambda_k
\end{aligned}$$

and the result follows. ∎

Remark 2.2.2 It can be shown that:

1. $E[\Lambda_1 \mid \mathcal{G}_0] = \Lambda_0 = 1$ where $\mathcal{G}_0 = \sigma(X_0)$.

2. $\Lambda_k \geq 0$ and $E[\Lambda_k] = \int_\Omega \Lambda_k dP = E[E[\Lambda_k \mid \mathcal{G}_{k-1}]] = \ \ldots \ = 1$. Therefore, Λ_k is a density.

□

Corollary 2.2.3 *$\overline{P}_k$ is a probability measure on $(\Omega, \mathcal{G}_k)$.*

Remark 2.2.4 It should be noted that:

1. The restriction of $\overline{P}_{k+1}$ to $\mathcal{G}_k$ is $\overline{P}_k$.

2. Let $\mathcal{G} = \bigvee_{k=0}^{\infty} \mathcal{G}_k$, the smallest σ-field generated by all the $\mathcal{G}_k$. The existence of $\overline{P}$ on $\mathcal{G}$ follows from Kolmogorov's Extension Theorem.

□

Definition 2.2.5 *On* $(\Omega, \mathcal{F}, P)$, *the probability measure* γ *is* absolutely continuous *with respect to probability measure* μ *if for each* A *in* $\mathcal{F}$, $\mu(A) = 0$ *implies* $\gamma(A) = 0$.

Theorem 2.2.6 *Suppose* $(\Omega, \mathcal{F}, \mathcal{P})$ *is a probability measure space,* $\mathcal{G} \subset \mathcal{F}$ *is a sub* σ*-field,* Q *is another probability measure absolutely continuous with respect to* P *and with Radon-Nikodym derivative* $\left(\frac{dQ}{dP}\right) = \Lambda$, *and* X *is any integrable* $\mathcal{F}$*-measurable random variable. Then:*

$$\overline{E}[X \mid \mathcal{G}] = \begin{cases} \frac{E[\Lambda X|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} & \text{if } E[\Lambda \mid \mathcal{G}] > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proof:

We must show for any set $A \in \mathcal{G}$:

$$\int_A \overline{E}[X \mid \mathcal{G}] d\overline{P} = \int_A (E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}]) d\overline{P}.$$

Let $G = \{\omega \in \Omega : E[\Lambda \mid \mathcal{G}] = 0\} \in \mathcal{G}$. Also, $G^c \in \mathcal{G}$.

Then $\int_G E[\Lambda \mid \mathcal{G}] dP = \int_G \Lambda dP = 0$.

Now $\Lambda \geq 0$ a.s., so either $\Lambda = 0$ a.s. on G or $P(G) = 0$. In either case, $\Lambda \equiv 0$ a.s. on G.

Consider any set $A \in \mathcal{G}$. Then $A = B \bigcup C$ where $B = A \bigcap G^c$ and $C = A \bigcap G$ and B and C are disjoint. Now $C \subset G$, so $\Lambda = 0$ a.s. on C and

$$\int_C \Lambda X dP = 0 = \int_C (E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}]) d\overline{P}.$$

Also:

$$\begin{aligned}
& \int_B (E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}])d\overline{P} \\
= \; & \overline{E}[I_B(E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}])] \\
= \; & E[\Lambda I_B(E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}])] \\
= \; & E[E[\Lambda I_B(E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}]) \mid \mathcal{G}]] \\
= \; & E[I_B(E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}])E[\Lambda \mid \mathcal{G}]] \\
= \; & E[E[I_B \Lambda X \mid \mathcal{G}]] \\
= \; & E[I_B \Lambda X] \\
= \; & \int_B \Lambda X dP.
\end{aligned}$$

Therefore, if $A = B \bigcup C$:

$$\int_A (E[\Lambda X \mid \mathcal{G}]/E[\Lambda \mid \mathcal{G}])d\overline{P} = \int_A \Lambda X dP = \int_A X d\overline{P} = \int_A \overline{E}[X \mid \mathcal{G}]d\overline{P}.$$

■

Note that under $\overline{P}$, the y process is a sequence of i.i.d. $N(0,1)$ random variables. In particular, the y_k are independent of the X_k.

Theorem 2.2.7 *Under $\overline{P}$, the y_k are i.i.d. $N(0,1)$ random variables.*

Proof: Now:

$$\begin{aligned}
& \overline{P}(y_k \leq t \mid \mathcal{G}_{k-1}) = \overline{E}[I(y_k \leq t) \mid \mathcal{G}_{k-1}] \\
= \; & \frac{E[\Lambda_k I(y_k \leq t) \mid \mathcal{G}_{k-1}]}{E[\Lambda_k \mid \mathcal{G}_{k-1}]} \quad (Theorem\ 2.2.5) \\
= \; & \frac{\Lambda_{k-1} E[\lambda_k I(y_k \leq t) \mid \mathcal{G}_{k-1}]}{\Lambda_{k-1} E[\lambda_k \mid \mathcal{G}_{k-1}]} .
\end{aligned}$$

Also:

$$\begin{aligned}
& E[\lambda_k \mid \mathcal{G}_{k-1}] = E\left[\frac{\gamma(X_{k-1})\phi(y_k)}{\phi(b_k)} \mid \mathcal{G}_{k-1}\right] \\
= \; & \gamma(X_{k-1})E\left[\frac{\phi(g(X_{k-1})+\gamma(X_{k-1})b_k)}{\phi(b_k)} \mid \mathcal{G}_{k-1}\right] \\
= \; & \gamma(X_{k-1})\int_{-\infty}^{\infty}\frac{\phi(g(X_{k-1})+\gamma(X_{k-1})b_k)}{\phi(b_k)}\phi(b_k)db_k \\
= \; & \gamma(X_{k-1})\int_{-\infty}^{\infty}\phi(z)\frac{dz}{\gamma(X_{k-1})} \quad (z = g(X_{k-1})+\gamma(X_{k-1})b_k) \\
= \; & 1.
\end{aligned}$$

Therefore:

$$\begin{aligned}
& P(y_k \le t \mid \mathcal{G}_{k-1}) \\
= \; & E[\lambda_k I(y_k \le t) \mid \mathcal{G}_{k-1}] \\
= \; & E\left[\frac{\gamma(X_{k-1})\phi(y_k)}{\phi(b_k)} I(y_k \le t) \mid \mathcal{G}_{k-1}\right] \\
= \; & \gamma(X_{k-1})\int_{-\infty}^{\infty}\frac{\phi(y_k)}{\phi(b_k)}\phi(b_k)I(y_k \le t)db_k \\
= \; & \int_{-\infty}^{\infty}\phi(y_k)I(y_k \le t)dy_k \\
= \; & \int_{-\infty}^{t}\phi(y_k)dy_k = \overline{P}(y_k \le t).
\end{aligned}$$

The result follows. ∎

Also note that under $\overline{P}$, the behaviour of the X process is unchanged; that is, X is still a Markov chain with transition matrix A and $X_{k+1} = AX_k + M_{k+1}$, where $\overline{E}[M_{k+1} \mid \mathcal{F}_k] = \underline{0}$.

Lemma 2.2.8 *Under $\overline{P}$, X is a Markov chain with the same transition probabilities as under P.*

Proof: Note that:

$$
\begin{aligned}
& \overline{P}(X_k = e_j \mid \mathcal{F}_{k-1}) = \overline{P}(X_k = e_j \mid \mathcal{G}_{k-1}) \\
= \; & \overline{E}[I(X_k = e_j) \mid \mathcal{G}_{k-1}] \\
= \; & \frac{\Lambda_{k-1} E[\lambda_k \langle X_k, e_j \rangle \mid \mathcal{G}_{k-1}]}{\Lambda_{k-1} E[\lambda_k \mid \mathcal{G}_{k-1}]} \\
= \; & E\left[\frac{\gamma(X_{k-1})\phi(y_k)}{\phi(b_k)} \langle X_k, e_j \rangle \mid \mathcal{G}_{k-1}\right] \\
= \; & \gamma(X_{k-1}) E\left[E\left[\frac{\phi(y_k)}{\phi(b_k)} \langle X_k, e_j \rangle \mid \mathcal{G}^+_{k-1}\right] \mathcal{G}_{k-1}\right] \quad (\mathcal{G}^+_{k-1} = \mathcal{G}_{k-1} \bigcup \{b_k\}) \\
= \; & \gamma(X_{k-1}) E\left[\frac{\phi(y_k)}{\phi(b_k)} E\left[\langle AX_{k-1} + M_k, e_j \rangle \mid \mathcal{G}^+_{k-1}\right] \mathcal{G}_{k-1}\right] \\
= \; & \langle AX_{k-1}, e_j \rangle E\left[\frac{\gamma(X_{k-1})\phi(y_k)}{\phi(b_k)} \mid \mathcal{G}_{k-1}\right] \\
= \; & \langle AX_{k-1}, e_j \rangle \; .
\end{aligned}
$$

Therefore, X is a Markov chain under $\overline{P}$ and $\overline{P}(X_k = e_j \mid \mathcal{F}_{k-1}) = \overline{P}(X_k = e_j \mid X_{k-1})$. If $X_{k-1} = e_i$, then $\overline{P}(X_k = e_j \mid X_{k-1} = e_i) = a_{ji}$ and the result follows. ∎

In light of the above observations, it is mathematically convenient to work under $\overline{P}$. However, P represents the "real-world" dynamics underlying our model. To construct P from $\overline{P}$ on $(\Omega, \mathcal{F})$, we define the b_k as $b_k := \frac{y_k - g(X_{k-1})}{\gamma(X_{k-1})}$, $k \in \mathbb{N}$, and probability measure P by $\left(\frac{dP}{d\overline{P}} \mid \mathcal{G}_k\right) = \overline{\Lambda}_k$, where $\overline{\Lambda}_k = \prod_{n=1}^{k} \overline{\lambda}_n$, $n \in \mathbb{N}$ and $n \geq 1$, $\overline{\Lambda}_0 = 1$, and $\overline{\lambda}_n = \frac{\phi(b_n)}{\gamma(X_{n-1})\phi(y_n)}$.

Lemma 2.2.9 *Under P, $\{b_k\}$, $k \in \mathbb{N}$, is a sequence of i.i.d. $N(0,1)$ random variables.*

Proof: The proof is similar to that of Theorem 2.2.6 .

Recursive Estimation:

We wish to estimate X_k, $k \in \mathbb{N}$, given $\mathcal{Y}_k$.

Lemma 2.2.10 *Suppose $X \in L^2(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$. Then the best estimate (in the mean square sense) in $L^2(\Omega, \mathcal{G}, P)$ of X is $E[X \mid \mathcal{G}]$.*

Proof: Suppose $Y \in L^2(\Omega, \mathcal{G}, P)$. Then:

$$\begin{aligned} & \| X - Y \|^2 = E[(X-Y)^2] \\ = \; & E[X^2] + E[E[Y^2 - 2XY \mid \mathcal{G}]] \\ = \; & E[X^2] + E[Y^2 - 2YE[X \mid \mathcal{G}]] \\ = \; & E[X^2] + E[Y - E(X \mid \mathcal{G})^2] - E[E(X \mid \mathcal{G})^2]. \end{aligned}$$

Therefore, $\| X - Y \|^2$ is minimized over integrable $\mathcal{G}$-measurable random variables when $Y = E[X \mid \mathcal{G}]$. ∎

It follows than that the best estimate (in L^2) of X_k given $\mathcal{Y}_k$ is $E[X_k \mid \mathcal{Y}_k]$. Earlier results indicate that:

$$E[X_k \mid \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]} \; .$$

Write $q_k = \overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k]$, $\Gamma^i(y_k) = \frac{\phi\left(\frac{y_k - a_i}{\gamma_i}\right)}{\gamma_i \phi(y_k)}$, and $a_i = Ae_i$.

Lemma 2.2.11 $\quad q_k = \sum_{i=1}^{N} \langle q_{k-1}, e_i \rangle \Gamma^i(y_k) a_i \; .$

Proof:

$$\begin{aligned} & q_k := \overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k] \\ = \; & \overline{E}[\overline{\Lambda}_{k-1} \overline{\lambda}_k (AX_{k-1} + M_k) \mid \mathcal{Y}_k] \end{aligned}$$

$$
\begin{aligned}
&= \overline{E}\left[\overline{\Lambda}_{k-1}\frac{\phi\left(\frac{y_k-g(X_{k-1})}{\gamma(X_{k-1})}\right)}{\gamma(X_{k-1})\phi(y_k)}(AX_{k-1}) \mid \mathcal{Y}_k\right] \\
&= \overline{E}\left[\sum_{i=1}^{N}\langle X_{k-1}, e_i\rangle\overline{\Lambda}_{k-1} \mid \mathcal{Y}_{k-1}\right]\Gamma^i(y_k)a_i \quad \left(\sum_{i=1}^{N}\langle X_{k-1}, e_i\rangle = 1\right) \\
&= \sum_{i=1}^{N}\langle\overline{E}[\overline{\Lambda}_{k-1}X_{k-1} \mid \mathcal{Y}_{k-1}], e_i\rangle\Gamma^i(y_k)a_i \ .
\end{aligned}
$$

The result follows. ∎

Lemma 2.2.12 $\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k] = \langle q_k, \underline{1}\rangle$.

Proof: Recall that $\sum_{i=1}^{N}\langle X_k, e_i\rangle = 1$.

Therefore, $\sum_{i=1}^{N}\langle\overline{\Lambda}_k X_k, e_i\rangle = \overline{\Lambda}_k$ and:

$$
\begin{aligned}
&\overline{E}\left[\sum_{i=1}^{N}\langle\overline{\Lambda}_k X_k, e_i\rangle \mid \mathcal{Y}_k\right] \\
&= \sum_{i=1}^{N}\langle\overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k], e_i\rangle \\
&= \sum_{i=1}^{N}\langle q_k, e_i\rangle = \overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k] \ .
\end{aligned}
$$

∎

In conclusion, we observe that $E[X_k \mid \mathcal{Y}_k] = \frac{q_k}{\langle q_k, \underline{1}\rangle}$, $k \in \mathcal{Z}^+$.

We also seek estimators for the occupation time in state r up to time $(k-1)$, the number of jumps from state r to state s in time k, and two processes related to observation drift and observation variance.

In what follows, let $J_k^r = \sum_{n=1}^{k}\langle X_{n-1}, e_r\rangle$, that is, the amount of time the X process has spent in state r up to time $(k-1)$; $N_k^{rs} = \sum_{n=1}^{k}\langle X_{n-1}, e_r\rangle\langle X_n, e_s\rangle$, that is, the number of jumps from state r to state s in time k; and $G_k^r(f) = \sum_{n=1}^{k}\langle X_{n-1}, e_r\rangle f(y_n)$ where f denotes either $f(y) = y$ or $f(y) = y^2$.

Remark 2.2.13 We observe the following:

1. Based on earlier results, $E[J_k^r \mid \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k J_k^r | \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k | \mathcal{Y}_k]}$ and $\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k] = \langle q_k, \underline{1} \rangle$.
2. $\sigma(J_k^r) := \overline{E}[\overline{\Lambda}_k J_k^r \mid \mathcal{Y}_k]$ does not provide a recurrence relation.
3. $\sigma(J_k^r X_k) := \overline{E}[\overline{\Lambda}_k J_k^r X_k \mid \mathcal{Y}_k]$ does provide a recurrence relation; (shown below).
4. Since $\sum_{i=1}^N \langle X_k, e_i \rangle = 1$, it follows that $\sum_{i=1}^N \langle \overline{\Lambda}_k J_k^r X_k, e_i \rangle = \overline{\Lambda}_k J_k^r$, $\sum_{i=1}^N \langle \overline{E}[\overline{\Lambda}_k J_k^r X_k \mid \mathcal{Y}_k], e_i \rangle = \overline{E}[\overline{\Lambda}_k J_k^r \mid \mathcal{Y}_k]$, and so $\sigma(J_k^r) = \langle \sigma(J_k^r X_k), \underline{1} \rangle$.
5. Therefore, $E[J_k^r \mid \mathcal{Y}_k] = \frac{\langle \sigma(J_k^r X_k), \underline{1} \rangle}{\langle q_k, \underline{1} \rangle}$.

□

Lemma 2.2.14 $\sigma(J_k^r X_k) = \sum_{i=1}^N \langle \sigma(J_{k-1}^r X_{k-1}), e_i \rangle \Gamma^i(y_k) a_i + \langle q_{k-1}, e_r \rangle \Gamma^r(y_k) a_r$.

Proof: Now:

$$\begin{aligned} J_k^r &= \sum_{n=1}^{k} \langle X_{n-1}, e_r \rangle \\ &= J_{k-1}^r + \langle X_{k-1}, e_r \rangle \end{aligned}$$

and:

$$J_k^r X_k = J_{k-1}^r X_k + \langle X_{k-1}, e_r \rangle X_k.$$

So:

$$\begin{aligned} & \sigma(J_k^r X_k) := \overline{E}[\overline{\Lambda}_k J_k^r X_k \mid \mathcal{Y}_k] \\ = \; & \overline{E}[\overline{\Lambda}_{k-1} \overline{\lambda}_k J_{k-1}^r (A X_{k-1} + M_k) \mid \mathcal{Y}_k] + \overline{E}[\overline{\Lambda}_{k-1} \overline{\lambda}_k \langle X_{k-1}, e_r \rangle (A X_{k-1} + M_k) \mid \mathcal{Y}_k] \\ = \; & \overline{E}\left[\sum_{i=1}^{N} \langle X_{k-1}, e_i \rangle \overline{\Lambda}_{k-1} J_{k-1}^r \mid \mathcal{Y}_{k-1} \right] \Gamma^i(y_k) a_i + \overline{E}[\overline{\Lambda}_{k-1} \langle X_{k-1}, e_r \rangle \mid \mathcal{Y}_{k-1}] \Gamma^r(y_k) a_r \\ = \; & \sum_{i=1}^{N} \langle \overline{E}[\overline{\Lambda}_{k-1} J_{k-1}^r X_{k-1} \mid \mathcal{Y}_{k-1}], e_i \rangle \Gamma^i(y_k) a_i + \langle \overline{E}[\overline{\Lambda}_{k-1} X_{k-1} \mid \mathcal{Y}_{k-1}], e_r \rangle \Gamma^r(y_k) a_r. \end{aligned}$$

The result follows. ■

Remark 2.2.15 Similar to the above, $E[N_k^{rs} \mid \mathcal{Y}_k] = \frac{\langle \sigma(N_k^{rs} X_k), \underline{1} \rangle}{\langle q_k, \underline{1} \rangle}$, where $\sigma(N_k^{rs} X_k) := \overline{E}[\overline{\Lambda}_k N_k^{rs} X_k \mid \mathcal{Y}_k]$.

□

Lemma 2.2.16 $\sigma(N_k^{rs} X_k) = \sum_{i=1}^{N} \langle \sigma(N_{k-1}^{rs} X_{k-1}), e_i \rangle \Gamma^i(y_k) a_i + \langle q_{k-1}, e_r \rangle \Gamma^r(y_k) a_{sr} e_s$.

Proof: Now:

$$\begin{aligned} N_k^{rs} &= \sum_{n=1}^{k} \langle X_{n-1}, e_r \rangle \langle X_n, e_s \rangle \\ &= N_{k-1}^{rs} + \langle X_{k-1}, e_r \rangle \langle X_k, e_s \rangle \end{aligned}$$

and:

$$N_k^{rs} X_k = N_{k-1}^{rs} X_k + \langle X_{k-1}, e_r \rangle \langle X_k, e_s \rangle e_s.$$

So:

$$\begin{aligned} & \sigma(N_k^{rs} X_k) := \overline{E}[\overline{\Lambda}_k N_k^{rs} X_k \mid \mathcal{Y}_k] \\ = \; & \overline{E}[\overline{\Lambda}_{k-1} \overline{\lambda}_k N_{k-1}^{rs} (A X_{k-1} + M_k) \mid \mathcal{Y}_k] + \\ & \qquad \overline{E}[\overline{\Lambda}_{k-1} \langle X_{k-1}, e_r \rangle \langle A X_{k-1} + M_k, e_s \rangle \mid \mathcal{Y}_k] \Gamma^r(y_k) e_s \\ = \; & \overline{E}\left[\sum_{i=1}^{N} \langle X_{k-1}, e_i \rangle \overline{\Lambda}_{k-1} N_{k-1}^{rs} \mid \mathcal{Y}_{k-1} \right] \Gamma^i(y_k) a_i + \\ & \qquad \overline{E}[\overline{\Lambda}_{k-1} \langle X_{k-1}, e_r \rangle \mid \mathcal{Y}_{k-1}] \Gamma^r(y_k) a_{sr} e_s \\ = \; & \sum_{i=1}^{N} \langle \overline{E}[\overline{\Lambda}_{k-1} N_{k-1}^{rs} X_{k-1} \mid \mathcal{Y}_{k-1}], e_i \rangle \Gamma^i(y_k) a_i + \\ & \qquad \langle \overline{E}[\overline{\Lambda}_{k-1} X_{k-1} \mid \mathcal{Y}_{k-1}], e_r \rangle \Gamma^r(y_k) a_{sr} e_s. \end{aligned}$$

The result follows. ■

Remark 2.2.17 $E[G_k^r(f) \mid \mathcal{Y}_k] = \frac{\langle \sigma(G_k^r(f)X_k), 1\rangle}{\langle q_k, 1\rangle}$, where $\sigma(G_k^r(f)X_k) := \overline{E}[\overline{\Lambda}_k G_k^r(f)X_k \mid \mathcal{Y}_k]$ and f denotes either $f(y) = y$ or $f(y) = y^2$.

□

Lemma 2.2.18 $\sigma(G_k^r(y)X_k) = \sum_{i=1}^N \langle \sigma(G_{k-1}^r(y)X_{k-1}), e_i\rangle \Gamma^i(y_k)a_i + \langle q_{k-1}, e_r\rangle y_k \Gamma^r(y_k)a_r$.

Proof: Now:

$$\begin{aligned} G_k^r(y) &= \sum_{n=1}^k \langle X_{n-1}, e_r\rangle y_n \\ &= G_{k-1}^r(y) + \langle X_{k-1}, e_r\rangle y_k \end{aligned}$$

and:

$$G_k^r(y)X_k = G_{k-1}^r(y)X_k + \langle X_{k-1}, e_r\rangle y_k X_k.$$

So:

$$\begin{aligned} & \sigma(G_k^r(y)X_k) := \overline{E}[\overline{\Lambda}_k G_k^r(y)X_k \mid \mathcal{Y}_k] \\ = \; & \overline{E}[\overline{\Lambda}_{k-1}\overline{\lambda}_k G_{k-1}^r(y)(AX_{k-1} + M_k) \mid \mathcal{Y}_k] + \\ & \qquad \overline{E}[\overline{\Lambda}_{k-1}\overline{\lambda}_k \langle X_{k-1}, e_r\rangle (AX_{k-1} + M_k) \mid \mathcal{Y}_k] y_k \\ = \; & \overline{E}\left[\sum_{i=1}^N \langle X_{k-1}, e_i\rangle \overline{\Lambda}_{k-1} G_{k-1}^r(y) \mid \mathcal{Y}_{k-1}\right] \Gamma^i(y_k)a_i + \\ & \qquad \overline{E}[\overline{\Lambda}_{k-1}\langle X_{k-1}, e_r\rangle \mid \mathcal{Y}_{k-1}] y_k \Gamma^r(y_k)a_r \\ = \; & \sum_{i=1}^N \langle \overline{E}[\overline{\Lambda}_{k-1} G_{k-1}^r(y)X_{k-1} \mid \mathcal{Y}_{k-1}], e_i\rangle \Gamma^i(y_k)a_i + \\ & \qquad \langle \overline{E}[\overline{\Lambda}_{k-1} X_{k-1} \mid \mathcal{Y}_{k-1}], e_r\rangle y_k \Gamma^i(y_k)a_r \; . \end{aligned}$$

The result follows. ■

Similarly, it can be shown that $\sigma(G_k^r(y^2)X_k) = \sum_{i=1}^N \langle \sigma(G_{k-1}^r(y^2)X_{k-1}), e_i\rangle \Gamma^i(y_k)a_i + \langle q_{k-1}, e_r\rangle y_k^2 \Gamma^r(y_k)a_r$.

Parameter Re-estimation:

Our model:

$$\begin{aligned} X_{k+1} &= AX_k + M_{k+1} \\ y_{k+1} &= g(X_k) + \gamma(X_k)b_{k+1} \\ &= \langle g, X_k \rangle + \langle \gamma, X_k \rangle b_{k+1}, \quad k \in \mathbb{N} - \end{aligned}$$

has $N^2 + 2N$ parameters which we seek to estimate based on the information in the y process. These parameters are the N^2 elements a_{ij}, $1 \leq i, j \leq N$, of the transition matrix A, the N components of the vector g, and the N components of the vector γ. The size of the state space of the Markov chain, which we have denoted by N, is not estimated; rather, we take it as given.

To estimate the above-mentioned parameters, we use the expectation maximization (EM) algorithm. [3] Based on the data, $y_1, \ldots, y_k$, we seek an estimate $\hat{\theta}$ of parameter θ that maximizes the likelihood function:

$$L(\theta) = E_{\theta_0}\left[\left(\frac{dP_\theta}{dP_{\theta_0}} \mid \mathcal{G}_k\right) \mid \mathcal{Y}_k\right].$$

Here, $\{P_\theta, \theta \in \Theta\}$ denotes a family of probability measures on measurable space $(\Omega, \mathcal{F})$ all absolutely continuous with respect to fixed probability measure P_{θ_0}; θ_0 denotes an initial or re-estimated value of θ; and $\mathcal{Y}_k \subset \mathcal{G}_k \subset \mathcal{F}$. In general, the maximum likelihood estimate is difficult to compute directly, and the EM algorithm provides an iterative approximation method:

1. Step 1. Set $p = 0$ and choose $\hat{\theta}_0$.

2. Step 2. (E-step) Set $\theta^* = \hat{\theta}_p$ and compute $Q(., \theta^*)$, where

$$Q(\theta, \theta^*) = E_{\theta^*}\left[\log\left(\frac{dP_\theta}{dP_{\theta^*}}\right) \mid \mathcal{Y}\right].$$

[3] The discussion of this paragraph follows Elliott, Aggoun, and Moore [8], pages 35-37.

3. Step 3. (M-step) Let $\hat{\theta}_{p+1}$ equal that value $\theta \in \Theta$ that maximizes $Q(\theta, \theta^*)$.

4. Step 4. Replace p by $p+1$ and repeat beginning with Step 2 until a stopping criterion is satisfied.

The EM algorithm implies that the estimates improve montonically; that is, the generated sequence $\{\hat{\theta}_p\}$ gives nondecreasing values of the likelihood function.

Lemma 2.2.19 *The value of g_i, $1 \leq i \leq N$, that maximizes the conditional likelihood is:*

$$\hat{g}_i = \frac{\sigma(G_k^i(y))}{\sigma(J_k^i)} \ .$$

Proof: The preceding analysis with respect to change of measure indicates that we have a density, $\overline{\Lambda}^g$, for each value of $g \in \mathbb{R}^N$.

Suppose we have an estimate for g and wish to use the data, $y_1, \ \ldots \ , y_k$, to obtain a better estimate, $\hat{g}$. To do this, we consider a change of measure from P_g to $P_{\hat{g}}$ given by:

$$\begin{aligned} \left(\frac{dP_{\hat{g}}}{dP_g} \mid \mathcal{G}_k\right) &= \frac{\overline{\Lambda}_k^{\hat{g}}}{\overline{\Lambda}_k^{g}} \\ &= \prod_{n=1}^{k} \frac{\overline{\lambda}_n^{\hat{g}}}{\overline{\lambda}_n^{g}} = \prod_{n=1}^{k} \hat{\lambda}_n. \end{aligned}$$

Now:

$$\begin{aligned} \hat{\lambda}_n &= \frac{\phi\left(\frac{y_n - \hat{g}(X_{n-1})}{\gamma(X_{n-1})}\right)}{\gamma(X_{n-1})\phi(y_n)} \Big/ \frac{\phi\left(\frac{y_n - g(X_{n-1})}{\gamma(X_{n-1})}\right)}{\gamma(X_{n-1})\phi(y_n)} \\ &= \exp\left(\frac{1}{2\gamma(X_{n-1})^2}\{g(X_{n-1})^2 - \hat{g}(X_{n-1})^2 - 2y_n g(X_{n-1}) + 2y_n \hat{g}(X_{n-1})\}\right). \end{aligned}$$

Then, we maximize $E\left[\log\left(\frac{dP_{\hat{g}}}{dP_g} \mid \mathcal{G}_k\right) \mid \mathcal{Y}_k\right]$ over $\hat{g}$, where E denotes expectation under g.

Now:

$$\begin{aligned}
&\log\left(\frac{dP_{\hat{g}}}{dP_g} \mid \mathcal{G}_k\right) \\
&= \sum_{n=1}^{k}\left(\frac{1}{2\gamma(X_{n-1})^2}\{g(X_{n-1})^2 - \hat{g}(X_{n-1})^2 - 2y_n g(X_{n-1}) + 2y_n\hat{g}(X_{n-1})\}\right) \\
&= \sum_{n=1}^{k}\sum_{i=1}^{N}\langle X_{n-1}, e_i\rangle\frac{1}{2\gamma_i^2}(g_i^2 - \hat{g}_i^2 - 2y_n g_i + 2y_n\hat{g}_i) \quad \left(\sum_{i=1}^{N}\langle X_{n-1}, e_i\rangle = 1\right) \\
&= \sum_{i=1}^{N}\left(\sum_{n=1}^{k}\langle X_{n-1}, e_i\rangle\frac{1}{2\gamma_i^2}(g_i^2 - \hat{g}_i^2 - 2y_n g_i + 2y_n\hat{g}_i)\right).
\end{aligned}$$

Also:

$$\begin{aligned}
&E\left[\log\left(\frac{dP_{\hat{g}}}{dP_g} \mid \mathcal{G}_k\right) \mid \mathcal{Y}_k\right] \\
&= \sum_{i=1}^{N}\frac{1}{2\gamma_i^2}\left(2\hat{g}_i E\left[\sum_{n=1}^{k}\langle X_{n-1}, e_i\rangle y_n \mid \mathcal{Y}_k\right] - \hat{g}_i^2 E\left[\sum_{n=1}^{k}\langle X_{n-1}, e_i\rangle \mid \mathcal{Y}_k\right]\right) + R(g) \\
&= \sum_{i=1}^{N}\frac{1}{2\gamma_i^2}(2\hat{g}_i E[G_k^i(y) \mid \mathcal{Y}_k] - \hat{g}_i^2 E[J_k^i \mid \mathcal{Y}_k]) + R(g)
\end{aligned}$$

where $R(g)$ involves only the g.

Differentiating in $\hat{g}_i$ and equating the derivative to zero, we see that the maximizing value of $\hat{g}_i$, given the observations up to time k, is:

$$\begin{aligned}
\hat{g}_i &= \frac{E[G_k^i(y) \mid \mathcal{Y}_k]}{E[J_k^i \mid \mathcal{Y}_k]} \\
&= \frac{\hat{E}[\hat{\Lambda}_k G_k^i(y) \mid \mathcal{Y}_k]/\hat{E}[\hat{\Lambda}_k \mid \mathcal{Y}_k]}{\hat{E}[\hat{\Lambda}_k J_k^i \mid \mathcal{Y}_k]/\hat{E}[\hat{\Lambda}_k \mid \mathcal{Y}_k]} \\
&= \frac{\sigma(G_k^i(y))}{\sigma(J_k^i)},
\end{aligned}$$

where $\hat{E}$ denotes expectation under $\hat{g}$. ∎

Lemma 2.2.20 *The value of γ_i, $1 \leq i \leq N$, that maximizes the conditional likelihood is:*

$$\hat{\gamma}_i = \frac{\sigma(G_k^i(y^2)) - 2g_i\sigma(G_k^i(y)) + g_i^2\sigma(J_k^i)}{\sigma(J_k^i)}.$$

Proof: The proof is similar to that of Lemma 2.2.18 .

In our model, the X process is a Markov chain with transition matrix A. We now wish to estimate A. We do this by using a density defined by:

$$\left(\frac{dP_{\hat{A}}}{dP_A} \mid \mathcal{G}_k\right) = \hat{\Lambda}_k = \prod_{n=1}^{k} \hat{\lambda}_n \ ,$$

where

$$\hat{\lambda}_n = \frac{\prod_{j=1}^{N} \langle \hat{A}X_{n-1}, e_j \rangle^{\langle X_n, e_j \rangle}}{\prod_{j=1}^{N} \langle AX_{n-1}, e_j \rangle^{\langle X_n, e_j \rangle}} \ .$$

It can be shown that, with $\overline{\Lambda}_k = \prod_{n=1}^{k} \overline{\lambda}_n$ and $\overline{\lambda}_n = N \prod_{j=1}^{N} \langle AX_{n-1}, e_j \rangle^{\langle X_n, e_j \rangle}$, X is a Markov chain with transition matrix A under P_A.

Given observations, $y_1, \ \ldots \ , y_k$, we wish to choose the transition matrix $\hat{A}$ that maximizes $E[\log \hat{\Lambda}_k \mid \mathcal{Y}_k]$ subject to the constraints that the columns of $\hat{A}$ have components whose values sum to one. Here, E denotes expectation under A.

Lemma 2.2.21 $\quad \hat{a}_{ji} = \frac{\sigma(N_k^{ij})}{\sigma(J_k^i)}$.

Proof:

$$\begin{aligned}
& E[\log \hat{\Lambda}_k \mid \mathcal{Y}_k] \\
= \ & E\left[\log \prod_{n=1}^{k} \prod_{j=1}^{N} \frac{\langle \hat{A}X_{n-1}, e_j \rangle^{\langle X_n, e_j \rangle}}{\langle AX_{n-1}, e_j \rangle^{\langle X_n, e_j \rangle}} \mid \mathcal{Y}_k\right] \\
= \ & E\left[\sum_{n=1}^{k} \sum_{j=1}^{N} \langle X_n, e_j \rangle \log(\langle \hat{A}X_{n-1}, e_j \rangle) + R(A) \mid \mathcal{Y}_k\right] \\
= \ & E\left[\sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{n=1}^{k} \langle X_{n-1}, e_i \rangle \langle X_n, e_j \rangle \log(\langle \hat{A}X_{n-1}, e_j \rangle) + R(A) \mid \mathcal{Y}_k\right] \\
= \ & E\left[\sum_{j=1}^{N} \sum_{i=1}^{N} N_k^{ij} \log(\hat{a}_{ji}) + R(A) \mid \mathcal{Y}_k\right]
\end{aligned}$$

where $R(A)$ depends only on A.

Now this is not a free choice for $\hat{a}_{ji}$; rather, we must have $\sum_{j=1}^{N}\langle \hat{A}X_n, e_j\rangle = 1$ or $\sum_{n=1}^{k}\sum_{j=1}^{N}\langle \hat{A}X_n, e_j\rangle = k$.

Using the Lagrange multiplier μ, we consider:

$$\begin{aligned} & F(\hat{a}_{11}, \hat{a}_{12}, \ \ldots \ , \hat{a}_{NN}, \mu) \\ = & \sum_{j=1}^{N}\sum_{i=1}^{N} E[N_k^{ij} \mid \mathcal{Y}_k]\log(\hat{a}_{ji}) + R(A) + \\ & \qquad \mu\left(E\left[\sum_{n=1}^{k}\sum_{j=1}^{N}\sum_{i=1}^{N}\langle X_n, e_i\rangle\langle \hat{A}X_n, e_j\rangle - k \mid \mathcal{Y}_k\right]\right) \\ = & \sum_{j=1}^{N}\sum_{i=1}^{N} E[N_k^{ij} \mid \mathcal{Y}_k]\log(\hat{a}_{ji}) + R(A) + \mu\left[\sum_{j=1}^{N}\sum_{i=1}^{N} E[J_k^i \mid \mathcal{Y}_k]\hat{a}_{ji} - k\right] . \end{aligned}$$

Differentiating F with respect to $\hat{a}_{ji}$, $1 \leq i, j \leq N$, and μ, in turn, and equating each derivative to zero, we get:

$$\begin{aligned} \hat{a}_{ji} &= \frac{E[N_k^{ij} \mid \mathcal{Y}_k]}{E[J_k^i \mid \mathcal{Y}_k]} \\ &= \frac{\hat{E}[\hat{\Lambda}_k N_k^{ij} \mid \mathcal{Y}_k]/\hat{E}[\hat{\Lambda}_k \mid \mathcal{Y}_k]}{E[\Lambda_k J_k^i \mid \mathcal{Y}_k]/E[\Lambda_k \mid \mathcal{Y}_k]} \\ &= \frac{\sigma(N_k^{ij})}{\sigma(J_k^i)} , \end{aligned}$$

where $\hat{E}$ denotes expectation under $\hat{A}$. ∎

2.2.2 Vector-Valued Observation Process

The observation process y may be vector-valued [4] with:

$$y_k = (y_k^1, \ \ldots \ , y_k^m), \quad k \in \mathcal{Z}^+ ;$$

[4] Important references in this section are Elliott [7], Section 7 and Elliott, Aggoun, and Moore [8], pages 70-72.

$$
\begin{aligned}
y_k^j &= \langle g^j, X_{k-1}\rangle + \langle \gamma^j, X_{k-1}\rangle b_k^j, \quad 1 \le j \le m\ ; \\
g^j &= (g_1^j, \ \ldots\ , g_N^j) \in \mathbb{R}^N, \quad 1 \le j \le m\ ; \quad \textit{and} \\
\gamma^j &= (\gamma_1^j, \ \ldots\ , \gamma_N^j) \in \mathbb{R}^N, \quad 1 \le j \le m.
\end{aligned}
$$

Here, the b_k^j, $1 \le j \le m$, $k \in \mathcal{Z}^+$, are i.i.d. $N(0,1)$ random variables.

We derive recursive relations similar to those given earlier:

$$
\begin{aligned}
q_k &= \sum_{i=1}^{N} \langle q_{k-1}, e_i\rangle \Gamma^i(y_k) a_i\ ; \\
\sigma(J_k^r X_k) &= \sum_{i=1}^{N} \langle \sigma(J_{k-1}^r X_{k-1}), e_r\rangle \Gamma^i(y_k) a_i + \langle q_{k-1}, e_r\rangle \Gamma^r(y_k) a_r\ ; \\
\sigma(N_k^{rs} X_k) &= \sum_{i=1}^{N} \langle \sigma(N_{k-1}^{rs} X_{k-1}), e_i\rangle \Gamma^i(y_k) a_i + \langle q_{k-1}, e_r\rangle \Gamma^r(y_k) a_{sr} e_s\ ; \quad \textit{and} \\
\sigma(G_k^r(y^j) X_k) &= \sum_{i=1}^{N} \langle \sigma(G_{k-1}^r(y^j) X_{k-1}), e_i\rangle \Gamma^i(y_k) a_i + \langle q_{k-1}, e_r\rangle y_k^j \Gamma^r(y_k) a_r
\end{aligned}
$$

where $k \in \mathcal{Z}^+$, $1 \le j \le m$, and:

$$
\begin{aligned}
\Gamma^i(y_k) &= \Gamma^i(y_k^1, \ \ldots\ , y_k^m) \\
&= \prod_{j=1}^{m} \frac{\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right)}{\gamma_i^j \phi(y_k^j)}\ , \quad 1 \le i \le N.
\end{aligned}
$$

Again, ϕ denotes the $N(0,1)$ density.

Based on the EM algorithm, the parameter estimators are:

$$
\begin{aligned}
\hat{g}_i^j &= \frac{\sigma(G_k^i(y^j))}{\sigma(J_k^i)}, 1 \le j \le m; \\
\hat{\gamma}_i^j &= \frac{\sigma(G_k^i((y^j)^2) - 2g_i^j \sigma(G_k^i(y^j)) + (g_i^j)^2 \sigma(J_k^i)}{\sigma(J_k^i)}, 1 \le j \le m; \ \textit{and} \\
\hat{a}_{ji} &= \frac{\sigma(N_k^{ij})}{\sigma(J_k^i)}, 1 \le j \le N.
\end{aligned}
$$

2.3 Zero Delay Observation Model

2.3.1 Real-Valued Observation Process

On the probability space $(\Omega, \mathcal{F}, P)$, we now define the zero delay observation model:[5]

$$\begin{aligned} X_k &= AX_{k-1} + M_k \\ y_k &= g(X_k) + \gamma(X_k)b_k,\ y_k \in \mathbb{R},\ k \in \mathcal{Z}^+ . \end{aligned}$$

Note that there is now no delay between the state X at time k and the observation y. Here, the process $\{b_k\}$, the functions g and γ, and the complete σ-fields $\mathcal{F}_k$, $\mathcal{Y}_k$, and $\mathcal{G}_k$ are defined as in section 2.2.1 .

Change of Measure:

To facilitate estimation of the state of the chain and the model parameters, based on the information in the y process, we again define a new probability measure $\overline{P}_k$ on $(\Omega, \mathcal{G}_k)$ by setting the restriction of the Radon-Nikodym derivative to $\mathcal{G}_k$ equal to Λ_k; that is, $\left(\frac{d\overline{P}_k}{dP} \mid \mathcal{G}_k\right) = \Lambda_k$ where now:

$$\begin{aligned} \Lambda_k &= \prod_{n=1}^{k} \lambda_n,\ n \geq 1; \\ \Lambda_0 &= 1; \quad \textit{and} \\ \lambda_n &= \frac{\langle \gamma, X_n \rangle \phi(y_n)}{\phi(b_n)}, \quad n \in \mathbb{N}. \end{aligned}$$

We then construct P from $\overline{P}$ by letting $\left(\frac{dP}{d\overline{P}} \mid \mathcal{G}_k\right) = \overline{\Lambda}_k$, where $\overline{\Lambda}_k = \prod_{n=1}^{k} \overline{\lambda}_n$, $n \geq 1$, and $\overline{\lambda}_n = \lambda_n^{-1}$.

[5] In this section, reference was made to Elliott(1994)[] and Elliott, Aggoun, and Moore(1995)[], pages 72-73.

Recursive Estimation:

We derive the following recursive relations that are similar, but not identical, to those of section 2.2.1 :

$$q_k = \sum_{i=1}^{N} \langle Aq_{k-1}, e_i \rangle \Gamma^i(y_k) e_i \tag{2.1}$$

$$\sigma(J_k^r X_k) = \sum_{i=1}^{N} \langle A\sigma(J_{k-1}^r X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i + \langle q_{k-1}, e_r \rangle \sum_{i=1}^{N} \Gamma^i(y_k) a_{ir} e_i \tag{2.2}$$

$$\sigma(N_k^{rs} X_k) = \sum_{i=1}^{N} \langle A\sigma(N_{k-1}^{rs} X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i + \langle q_{k-1}, e_r \rangle \Gamma^s(y_k) a_{sr} e_s \quad \textit{and}$$

$$\sigma(G_k^r(f_k) X_k) = \sum_{i=1}^{N} \langle A\sigma(G_{k-1}^r(f_{k-1}) X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i$$

$$+ \langle q_{k-1}, e_r \rangle f_k \sum_{i=1}^{N} \Gamma^i(y_k) a_{ir} e_i. \tag{2.3}$$

In (2.3), $f_k = y_k$ or $f_k = y_k^2$. We give proofs for (2.1) and (2.2).

Proof: (of (2.1))

$$\begin{aligned}
q_k &:= \overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k] \\
&= \overline{E}\left[\sum_{i=1}^{N} \langle X_k, e_i \rangle \overline{\Lambda}_{k-1} \frac{\phi\left(\frac{y_k - a_i}{\sigma_i}\right)}{\sigma_i \phi(y_k)} e_i \mid \mathcal{Y}_k\right] \\
&= \overline{E}\left[\sum_{i=1}^{N} \langle AX_{k-1} + M_k, e_i \rangle \overline{\Lambda}_{k-1} \mid \mathcal{Y}_k\right] \Gamma^i(y_k) e_i \\
&= \sum_{i=1}^{N} \overline{E}[\langle AX_{k-1}, e_i \rangle \overline{\Lambda}_{k-1} \mid \mathcal{Y}_{k-1}] \Gamma^i(y_k) e_i \\
&= \sum_{i=1}^{N} \langle A\overline{E}[\overline{\Lambda}_{k-1} X_{k-1} \mid \mathcal{Y}_{k-1}], e_i \rangle \Gamma^i(y_k) e_i \\
&= \sum_{i=1}^{N} \langle Aq_{k-1}, e_i \rangle \Gamma^i(y_k) e_i.
\end{aligned}$$

■

Proof: (of (2.2))

$$
\begin{aligned}
& \sigma(J_k^r X_k) := \overline{E}[\overline{\Lambda}_k J_k^r X_k \mid \mathcal{Y}_k] \\
= \; & \overline{E}[\overline{\Lambda}_{k-1}\overline{\lambda}_k J_{k-1}^r X_k \mid \mathcal{Y}_k] + \overline{E}[\overline{\Lambda}_{k-1}\overline{\lambda}_k \langle X_{k-1}, e_r\rangle X_k \mid \mathcal{Y}_k] \\
= \; & \overline{E}\left[\sum_{i=1}^{N} \langle X_k, e_i\rangle \overline{\Lambda}_{k-1} J_{k-1}^r \mid \mathcal{Y}_k\right] \Gamma^i(y_k) e_i \\
& \qquad + \overline{E}\left[\sum_{i=1}^{N} \langle X_n, e_i\rangle \overline{\Lambda}_{k-1} \langle X_{k-1}, e_r\rangle \mid \mathcal{Y}_k\right] \Gamma^i(y_k) e_i \\
= \; & \sum_{i=1}^{N} \overline{E}[\langle AX_{k-1} + M_k, e_i\rangle \overline{\Lambda}_{k-1} J_{k-1}^r \mid \mathcal{Y}_k] \Gamma^i(y_k) e_i \\
& \qquad + \sum_{i=1}^{N} \overline{E}[\langle AX_{k-1} + M_k, e_i\rangle \overline{\Lambda}_{k-1} \langle X_{k-1}, e_r\rangle \mid \mathcal{Y}_k] \Gamma^i(y_k) e_i \\
= \; & \sum_{i=1}^{N} \langle A\overline{E}[\overline{\Lambda}_{k-1} J_{k-1}^r X_{k-1} \mid \mathcal{Y}_{k-1}], e_i\rangle \Gamma^i(y_k) e_i \\
& \qquad + \sum_{i=1}^{N} \overline{E}[\overline{\Lambda}_{k-1} \langle AX_{k-1}, e_i\rangle \langle X_{k-1}, e_r\rangle \mid \mathcal{Y}_{k-1}] \Gamma^i(y_k) e_i \\
= \; & \sum_{i=1}^{N} \langle A\sigma(J_{k-1}^r X_{k-1}), e_i\rangle \Gamma^i(y_k) e_i \\
& \qquad + \sum_{i=1}^{N} \langle \overline{E}[\overline{\Lambda}_{k-1} X_{k-1} \mid \mathcal{Y}_{k-1}], e_r\rangle \Gamma^i(y_k) a_{ir} e_i
\end{aligned}
$$

and the result follows. ∎

Parameter Re-estimation:

Using the EM algorithm, we derive the following parameter estimators:

$$
\begin{aligned}
\hat{g}_i &= \frac{\sigma(G_k^i(y))}{\sigma(J_k^i)}, 1 \le i \le N; \\
\hat{\gamma}_i &= \frac{\sigma(G_k^i(y^2)) - 2g_i\sigma(G_k^i(y)) + g_i^2\sigma(J_k^i)}{\sigma(J_k^i)}, 1 \le i \le N; \text{ and} \\
\hat{a}_{ji} &= \frac{\sigma(N_k^{ij})}{\sigma(J_k^i)}, 1 \le i, j \le N.
\end{aligned}
$$

2.3.2 Vector-Valued Observation Process

In the zero delay observation process, the y process may also be vector-valued with:

$$
\begin{aligned}
& y_k = (y_k^1, \ \ldots \ , y_k^m), \ k \in \mathcal{Z}^+; \\
& y_k^j = \langle g^j, X_k \rangle + \langle \gamma^j, X_k \rangle b_k^j, \ 1 \leq j \leq m; \ \textit{and} \\
& g^j, \gamma^j \in \mathbb{R}^N.
\end{aligned}
$$

The $b_k^j, 1 \leq j \leq m, k \in \mathcal{Z}^+$, are i.i.d. $N(0,1)$ random variables.

The following recursive relations can be derived:

$$
\begin{aligned}
q_k &= \sum_{i=1}^{N} \langle A q_{k-1}, e_i \rangle \Gamma^i(y_k) e_i \ ; \\
\sigma(J_k^r X_k) &= \sum_{i=1}^{N} \langle A \sigma(J_{k-1}^r X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i \\
&\qquad + \langle q_{k-1}, e_r \rangle \sum_{i=1}^{N} \Gamma^i(y_k) a_{ir} e_i \ ; \\
\sigma(N_k^{rs} X_k) &= \sum_{i=1}^{N} \langle A \sigma(N_{k-1}^{rs} X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i \\
&\qquad + \langle q_{k-1}, e_r \rangle \Gamma^s(y_k) a_{sr} e_s \ ; \ \ \textit{and} \\
\sigma(G_k^r(y^j) X_k) &= \sum_{i=1}^{N} \langle A \sigma(G_{k-1}^r(y^j) X_{k-1}), e_i \rangle \Gamma^i(y_k) e_i \\
&\qquad + \langle q_{k-1}, e_r \rangle y_k^j \sum_{i=1}^{N} \Gamma^i(y_k) a_{ir} e_i
\end{aligned}
$$

where $k \in \mathcal{Z}^+$, $1 \leq j \leq m$, and

$$
\begin{aligned}
\Gamma^i(y_k) &= \Gamma^i(y_k^1, \ \ldots \ , y_k^m) \\
&= \prod_{j=1}^{m} \frac{\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right)}{\gamma_i^j \phi(y_k^j)}, \ 1 \leq i \leq N.
\end{aligned}
$$

Again, ϕ denotes the $N(0,1)$ density.

Using the EM algorithm, we get parameter estimators, $\hat{g}_i^j$, $\hat{\gamma}_i^j$, and $\hat{a}_{ji}$, $1 \leq i, j \leq m$, identical to those of section 2.2.2 .

Chapter 3

Application: Asset Pricing

This chapter applies the theory of Chapter 2 to the problem of asset pricing. [1] It makes some preliminary observations and then discusses the model, price prediction, and applications and results.

3.1 Preliminaries

The application in this chapter is motivated by the log-normal model for price processes in continuous time. In a 1965 paper, Paul Samuelson [19] proposed that the price of a risky asset might evolve according to the following stochastic differential equation:

$$\frac{dS_t}{S_t} = \mu dt + \gamma dw_t, \; t \geq 0 \tag{3.1}$$

where S_t denotes the price of an asset at time t; μ and γ are model parameters (assumed constant here); and w_t represents Brownian motion.

Using the Ito formula which arises from the Ito calculus for stochastic integrals [2], it can be demonstrated that equation (3.1) has solution:

[1] Important references in this chapter are Elliott, Aggoun, and Moore [8], Elliott and Hunter [9], and Elliott, Hunter, and Jamieson [10].

[2] Stochastic integrals are discussed in, for example, Oksendal [17].

$$S_t = S_0 \exp\left(\left(\mu - \frac{\gamma^2}{2}\right) t + \gamma w_t\right)$$

or

$$\log\left(\frac{S_t}{S_0}\right) = \left(\mu - \frac{\gamma^2}{2}\right) t + \gamma w_t \; ; \; t \geq 0.$$

In fact:

$$\log\left(\frac{S_{t+h}}{S_t}\right) = \left(\mu - \frac{\gamma^2}{2}\right) h + \gamma\left(w_{t+h} - w_t\right); \; t, h \geq 0 \; . \tag{3.2}$$

3.2 The Model

We now turn to a discrete time price process $\{S_k; k \in \mathcal{Z}^+\}$. Motivated by (3.2), we propose a model of the following form:

$$\log\left(\frac{S_{k+1}}{S_k}\right) = g_k + \gamma_k b_{k+1} \; .$$

Here, the b_{k+1} are i.i.d. $N(0,1)$ random variables. The functions g_k and γ_k now depend on k and play the roles of $\left(\mu - \frac{\gamma^2}{2}\right)$ and γ, respectively. In fact, we suppose there is some discrete time, finite state Markov chain X such that $g_k = g\left(X_k\right)$ and $\gamma_k = \gamma\left(X_k\right)$. The state space of X is taken to be the set of unit vectors in $\mathbb{R}^N$, where N denotes the cardinality of the state space of X. Then, as in Chapter 2:

$$\begin{aligned} X_{k+1} &= AX_k + M_{k+1} && (3.3) \\ y_{k+1} &= \log\left(\frac{S_{k+1}}{S_k}\right) \\ &= g_k + \gamma_k b_{k+1} \\ &= g(X_k) + \gamma(X_k) b_{k+1} \\ &= \langle g, X_k \rangle + \langle \gamma, X_k \rangle b_{k+1} \; , \; k \in \mathbb{N} \; , && (3.4) \end{aligned}$$

where S_k denotes the price of an asset at time k. This is the unit delay observation model of Chapter 2; furthermore, the observation process, $y = \{y_k\}$, is real-valued. Thus all the results of Chapter 2, section 2.2.1 carry over to the discussion here.

In this model, g_k represents the drift in prices at time k and γ_k the volatility. The Markov process $X = \{X_k\}$ drives the evolution of these parameters and hence it drives the evolution of asset prices.

Process X is not observed directly in the model; rather, process y gives noisy observations of X. From the y_k observations, we seek to estimate the state of the chain at a given time, the transition probabilities associated with the chain, and the components of the vectors g and γ. We also seek to make price predictions.

The estimators for the parameters enumerated above are given in Chapter 2, section 2.2.1 .

3.3 Price Prediction

Based on the information in the observed asset prices, using the results of Chapter 2 we seek to predict future prices. In what follows, we give both the one-step-ahead price predictor and the two-step-ahead predictor.

Lemma 3.3.1 One-step-ahead predictor:

$$E[S_{k+1} \mid \mathcal{Y}_k] = S_k \sum_{i=1}^{N} e^{g_i + \frac{\gamma_i^2}{2}} \frac{\langle q_k, e_i \rangle}{\langle q_k, \underline{1} \rangle} \ .$$

Proof:

$$\begin{aligned} & E[S_{k+1} \mid \mathcal{Y}_k] \\ = \ & E\left[S_k \exp\left(g(X_k) + \gamma(X_k) b_{k+1}\right) \mid \mathcal{Y}_k\right] \end{aligned}$$

$$
\begin{aligned}
&= S_k \, E\left[\sum_{i=1}^{N} \langle X_k, e_i \rangle \exp\left(g(X_k) + \gamma(X_k) b_{k+1}\right) \mid \mathcal{Y}_k\right] \\
&= S_k \sum_{i=1}^{N} E\left[\langle X_k, e_i \rangle e^{g_i + \gamma_i b_{k+1}} \mid \mathcal{Y}_k\right] \\
&= S_k \sum_{i=1}^{N} e^{g_i} \frac{\langle q_k, e_i \rangle}{\langle q_k, \underline{1} \rangle} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\gamma_i b} e^{-\frac{b^2}{2}} db \; .
\end{aligned}
$$

By completing the square:

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\gamma_i b - \frac{b^2}{2}} db \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(b-\gamma_i)^2} e^{\frac{\gamma_i^2}{2}} db \\
&= e^{\frac{\gamma_i^2}{2}} \; .
\end{aligned}
$$

The result follows. ∎

Lemma 3.3.2 Two-step-ahead predictor:

$$
E[S_{k+2} \mid \mathcal{Y}_k] = S_k \sum_{i=1}^{N} \sum_{j=1}^{N} e^{g_i + \frac{\gamma_i^2}{2}} \frac{\langle q_k, e_i \rangle}{\langle q_k, \underline{1} \rangle} e^{g_j + \frac{\gamma_j^2}{2}} a_{ji} \; .
$$

Proof:

$$
\begin{aligned}
S_{k+1} &= S_k \exp\left(g(X_k) + \gamma(X_k) b_{k+1}\right) \\
&= S_k \sum_{i=1}^{N} \langle X_k, e_i \rangle \exp\left(g_i + \gamma_i b_{k+1}\right) \; .
\end{aligned}
$$

From lemma 3.3.1:

$$
E[S_{k+1} \mid \mathcal{G}_k] = S_k \sum_{i=1}^{N} \langle X_k, e_i \rangle e^{g_i + \frac{\gamma_i^2}{2}} \; .
$$

Therefore:

$$
E[S_{k+2} \mid \mathcal{G}_{k+1}] = S_{k+1} \sum_{j=1}^{N} \langle X_{k+1}, e_j \rangle e^{g_j + \frac{\gamma_j^2}{2}}
$$

and:

$$\begin{aligned} & E[E[S_{k+2} \mid \mathcal{G}_{k+1}]\mathcal{G}_k] \\ = \; & S_k \sum_{i=1}^{N} \langle X_k, e_i \rangle e^{g_i + \frac{\gamma_i^2}{2}} \sum_{j=1}^{N} \langle AX_k + M_{k+1}, e_j \rangle e^{g_j + \frac{\gamma_j^2}{2}} \\ = \; & S_k \sum_{i=1}^{N} \sum_{j=1}^{N} e^{g_i + \frac{\gamma_i^2}{2}} \langle X_k, e_i \rangle e^{g_j + \frac{\gamma_j^2}{2}} a_{ji} \; . \end{aligned}$$

It follows that:

$$E[S_{k+2} \mid \mathcal{Y}_k] = S_k \sum_{i=1}^{N} \sum_{j=1}^{N} e^{g_i + \frac{\gamma_i^2}{2}} \frac{\langle q_k, e_i \rangle}{\langle q_k, \underline{1} \rangle} e^{g_j + \frac{\gamma_j^2}{2}} a_{ji} \; .$$

■

3.4 Applications and Results

This section discusses three applications in asset pricing. It begins with some general comments and observations applicable to all three cases and then considers each case in turn.

3.4.1 The Program

A program was written in the **S** language to implement the estimation procedure of Chapter 2, section 2.2.1 . A copy of the program is given in Appendix A.

The program uses an iterative procedure. First, it assigns starting values to the model parameters. Then taking at least two prices, it estimates parameter values. With these estimates and at least one further observation, parameter values are re-estimated. The procedure can be repeated any number of times. As mentioned in Chapter 2, the EM algorithm implies that parameter estimates improve monotonically with each iteration.

The only parameter value that is not estimated is N, the size of the state space of the Markov chain X. In choosing a value for N, we were guided by the following considerations: on the one hand, we wanted N large enough that we could calculate $E[J_k^r \mid \mathcal{Y}_k]$ and $E[N_k^{rs} \mid \mathcal{Y}_k]$ and on the other hand, we sought a parsimonious model. In the three cases discussed below, N was assigned a value of four.

3.4.2 Starting Values

Starting values were assigned to the vectors g and γ and the matrix A. A number of approaches were taken to obtain starting values.

First, guided by the continuous time model, we determined starting values for the g and γ vectors in the following way:

1. Choose $m, n \in \mathcal{Z}^+$.
2. Take $N = mn$.
3. Any i, $1 \leq i \leq N$, has unique representation: $i = (q-1)m + p$ for $1 \leq p \leq m$, $1 \leq q \leq n$. This gives vectors p and q in $\mathbb{R}^N$.
4. Take $g_i = \frac{m-2p_i}{2N}$ and $\gamma_i = \frac{q_i}{N}$, $1 \leq i \leq N$.

All entries in the A matrix were assigned a starting value of $\frac{1}{N}$.

Second, based upon consideration of the simulations in Elliott [7] and Krishnamurthy and Moore [14] and assuming $N = 4$, we tried the following starting values: $g = (1.0, 0.0, 1.0, 0.0)'$, $\gamma = (0.25, 0.25, 0.25, 0.25)'$, and $a_{ij} = 0.25$, $1 \leq i, j \leq 4$. We also tried minor variations of these starting values.

This subject is discussed further in the three case studies.

3.4.3 The Time Step

The price data that were analyzed were observed at varying time intervals, for example, weekly or monthly. To incorporate into the analysis the fact that the time interval between observations can vary, we redefined the ϕ function of Chapter 2 to be the $N(0, dt)$ density, where dt represents the time step between observations. The discussion of Chapter 2 had assumed $dt = 1$.

Now, the model is:

$$\begin{aligned} X_{k(dt)} &= AX_{(k-1)(dt)} + M_{k(dt)} \\ y_{k(dt)} &= \log\left(\frac{S_{k(dt)}}{S_{(k-1)(dt)}}\right) \\ &= g(X_k) + \tilde{\gamma}(X_k) b_{k(dt)} \\ &= g(X_k) + \gamma(X_k)\frac{b_{k(dt)}}{\sqrt{dt}} \\ &= \langle g, X_k\rangle + \langle \gamma, X_k\rangle \frac{b_{k(dt)}}{\sqrt{dt}} \ , \end{aligned}$$

where $k \in \mathcal{Z}^+$, $dt \in \mathbb{R}^+$, and $M_{k(dt)}$ is a martingale increment. Here, $b_{k(dt)}$ is distributed $N(0, dt)$ and $\frac{b_{k(dt)}}{\sqrt{dt}}$ is distributed $N(0, 1)$. Also, $\gamma, \tilde{\gamma} \in \mathbb{R}^N$ and $\tilde{\gamma}_i = \frac{\gamma_i}{\sqrt{dt}}$.

In what follows, we assume that the noise in the observations is $N(0, 1)$. Thus, we estimate the vectors g and γ and then obtain the vector of interest, $\tilde{\gamma}$, by setting $\tilde{\gamma}_i = \frac{\gamma_i}{\sqrt{dt}}$, $1 \leq i \leq N$. Also, we revert to our earlier representation of the model, as in equations (3.3) and (3.4), since, assuming $N(0, 1)$ noise, the two representations are equivalent if we simply define the time interval between observations to be one unit of time.

The time step is measured in number of days between observations *or* number of working days between observations divided by 252, 252 being (approximately) the total number of working days in a year.

3.4.4 Checking the Model

The equation:

$$y_{k+1} = g(X_k) + \gamma(X_k)b_{k+1} \ ,$$

where $k \in \mathbb{N}$ and $\{b_{k+1}\}$ is a sequence of i.i.d. $N(0,1)$ random variables, suggests that the statistic:

$$\frac{\log\left(\frac{S_{k+1}}{S_k}\right) - g(\hat{X}_k)}{\gamma(\hat{X}_k)},$$

where $\hat{X}_k = \frac{q_k}{\langle q_k, \underline{1} \rangle}$, should be normally distributed with mean zero and variance one. In the case studies that follow, we check the model by testing for the normality of this statistic and by calculating its mean and variance.

Also, because N_k^{rs} represents the number of jumps from state r to state s up to time k, the value of $E[N_k^{rs} \mid \mathcal{Y}_k] = \frac{\langle \sigma(N_k^{rs} X_k), \underline{1} \rangle}{\langle q_k, \underline{1} \rangle}$, $k \in \mathcal{Z}^+$, must be less than the number of time steps up to time k. We check this in the case studies below.

3.4.5 Magnifying the Price Changes

It was observed that $\Gamma^i(y_k) = \frac{\phi\left(\frac{y_k - g_i}{\gamma_i}\right)}{\gamma_i \phi(y_k)}$, $1 \leq i \leq N$, $1 \leq k \leq K$, goes to zero for small γ_i and large $|y_k - g_i|$. If this occurs, parameter estimates are not obtained. To avoid this, we specified a tolerance value, usually $1.0e$-10, below which we would not let the value of $\Gamma^i(y_k)$ fall. However, if the values of y_k, g_i, and γ_i were such that $\Gamma^i(y_k)$ assumed this tolerance value for most values of i and k, important price information was lost to the subsequent parameter estimation.

We found it useful to multiply the y_k values by a factor $\alpha > 1$. By magnifying the price changes in this way, we obtained larger $\Gamma^i(y_k)$ values. However, having multiplied the y_k by $\alpha > 1$, we had, of course, to adjust the parameter estimates and price predictions that were subsequently derived.

If we consider the model:

$$\begin{aligned} y_k &= \log\left(\frac{S_k}{S_{k-1}}\right) \\ &= g(X_{k-1}) + \gamma(X_{k-1})b_k \\ &= \langle g, X_{k-1}\rangle + \langle \gamma, X_{k-1}\rangle b_k - \end{aligned}$$

where $k \in \mathcal{Z}^+$ and $\{b_k\}$ is a sequence of i.i.d. $N(0,1)$ random variables, then:

$$\begin{aligned} & Y_k = \alpha y_k \\ = \;& \alpha\langle g, X_{k-1}\rangle + \alpha\langle \gamma, X_{k-1}\rangle b_k \\ = \;& \langle \alpha g, X_{k-1}\rangle + \langle \alpha\gamma, X_{k-1}\rangle b_k \\ = \;& \langle g(\alpha), X_{k-1}\rangle + \langle \gamma(\alpha), X_{k-1}\rangle b_k - \end{aligned}$$

where $g(\alpha) = \alpha g$ and $\gamma(\alpha) = \alpha\gamma$. Therefore, $\hat{g(\alpha)} = \alpha\hat{g}$ and $\hat{g} = \frac{\hat{g(\alpha)}}{\alpha}$. Similarly, $\hat{\gamma} = \frac{\hat{\gamma(\alpha)}}{\alpha}$.

Also:

$$\begin{aligned} & E[S_{k+1} \mid \mathcal{Y}_k] \\ = \;& S_k\, E\left[\exp\left(\frac{1}{\alpha}Y_{k+1}\right) \mid \mathcal{Y}_k\right] \\ = \;& S_k\, E\left[\exp\left(\frac{1}{\alpha}\left(\langle g(\alpha), X_k\rangle + \langle \gamma(\alpha), X_k\rangle b_{k+1}\right)\right) \mid \mathcal{Y}_k\right] \\ = \;& S_k\, E\left[\sum_{i=1}^{N}\langle X_k, e_i\rangle \exp\left(\langle \frac{g(\alpha)}{\alpha}, X_k\rangle + \langle \frac{\gamma(\alpha)}{\alpha}, X_k\rangle b_{k+1}\right) \mid \mathcal{Y}_k\right] \\ = \;& S_k \sum_{i=1}^{N} E\left[\langle X_k, e_i\rangle \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i(\alpha)}{\alpha}b_{k+1}\right) \mid \mathcal{Y}_k\right] \\ = \;& S_k \sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha}\right) \frac{\langle q_k, e_i\rangle}{\langle q_k, \underline{1}\rangle} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{\gamma_i(\alpha)}{\alpha}b - \frac{b^2}{2}}\, db \\ = \;& S_k \sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right) \frac{\langle q_k, e_i\rangle}{\langle q_k, \underline{1}\rangle}\,. \end{aligned} \tag{3.5}$$

3.4.6 Case 1: IBM Stock Prices

In this first case, we analyzed a data set consisting of 240 monthly observations on the price of IBM stock. The sample period ran from November 1975 to October 1995. The data were compiled by Star Data Systems Inc. and published in *The Financial Post.*

The data were processed in 15 groups of 16 prices each; at the end of each pass through the data, parameter estimates were updated using the formulas of Chapter 2, section 2.2.1 .

Size of N:

As noted earlier, in the analysis the size of the state space of the Markov chain, denoted by N, was taken to be four. We did experiment with other values of N, for example, N= 2, 9, 16, and 25 and compared the results using $\hat{g}_{15}$ and $\hat{\gamma}_{15}$, computed after the fifteenth pass through the data. Note that:

$$\hat{g}_k = E[\langle g, X_k \rangle] = \langle g, \hat{X}_k \rangle \tag{3.6}$$

and

$$\hat{\gamma}_k = E[\langle \gamma, X_k \rangle] = \langle \gamma, \hat{X}_k \rangle \tag{3.7}$$

where

$$\hat{X}_k = E[X_k \mid \mathcal{Y}_k] = \frac{q_k}{\langle q_k, \underline{1} \rangle} \ , \ 1 \leq k \leq K.$$

Here, K=15 since $y_k, 1 \leq k \leq K$, is observed 15 times in each pass. Table 3.1 gives the values of $\hat{g}_{15}$ and $\hat{\gamma}_{15}$, after the fifteenth pass, for the cases N= 2, 4, and 9. It can be observed that these values are of the same order of magnitude across the three cases.

The Time Step:

Because the data consisted of monthly observations, we used a time step of 21, 21 being the number of working days between observations. As an alternative, we could have used a time step of 21/252. However, we found, using this latter value, that $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$, which is calculated in each of the 15 passes, assumed the tolerance value of 1.0e-10 785 out of the 900 possible times. As a result, information contained in the prices was lost to subsequent parameter estimation.

Starting Values:

The starting values that were assigned to the distribution of the state of the Markov chain, that is, to $E[X_0]$, the transition matrix A, and the vectors g and γ are reported in Table 3.2 . We also did the analysis using the starting values based on the continuous time model (discussed in section 3.4.2). In this latter case, the components of the estimated g and γ vectors were small; for example, the estimated γ vector, after the fifteenth pass, was $(0.0004802504, 0.0136553556, 0.0002319098, 0.0047173382)'$. Thus, we rejected these starting values in favour of the former since a problem with our results appears to be an inability of the model to capture fully the variability in the data. This is discussed further in what follows.

Magnifying the Price Changes and Checking the Model:

In the analysis, we multiplied the y_k values, $1 \leq k \leq 15$, by a factor of $\alpha = 2.5$ in each pass through the data. We did the analysis using other values of α as well, for example, α= 1.0, 5.0, 7.5, 10.0, 20.0, and 100.0 and then selected that

value of α that gave the best results based on the following considerations: the values of $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$ in each of the fifteen passes; the price prediction results; and the checks on the model (discussed in section 3.4.4). For example, using $\alpha = 2.5$ rather than $\alpha = 1.0$, we were able to reduce the number of times $\Gamma^i(y_k)$ assumed the tolerance value of 1.0e-10 from 371 to 183, out of a possible total of 900. Also, in regard to our two checks on the model, we report, in Table 3.3, the mean and variance of the test statistic:

$$\frac{\log\left(\frac{S_k}{S_{k-1}}\right) - g(\hat{X}_{k-1})}{\gamma(\hat{X}_{k-1})} \ , \ 1 \leq k \leq 15 \ -$$

and the correlation between this statistic and its normal scores, calculated for each pass through the data. These results are reported for the case $\alpha = 2.5$. As was mentioned in section 3.4.4, this statistic should be normally distributed with a mean of zero and a variance of one. The results of Table 3.3 support the hypothesis of normality, based on the Ryan-Joiner test [16], as well as the hypothesis of zero mean.[3] However, the hypothesis of unit variance is not supported. These results are, however, superior to those obtained for other values of α. The second check on the model, based on the value of $E[N_k^{r^s} \mid \mathcal{Y}_k]$, is also supported for the case $\alpha = 2.5$.

Parameter Re-estimation:

Table 3.2 gives the re-estimated values of the conditional distribution of the state of the Markov chain, that is, of $E[X_{15} \mid \mathcal{Y}_{15}]$, the transition matrix A, and the vectors g and γ, after the fifteenth pass through the data. These results are for the case $\alpha = 2.5$.

[3]The 95 percent confidence interval for the sample mean is $(-0.790, 3.240)$.

Price Prediction:

At the end of each pass through the data, we calculated the predicted price for the following month using formula (3.5). For example, the first pass uses the prices at the beginning of each month for the period November 1, 1975 to February 1, 1977 and predicts the price for March 1, 1977. Proceeding in this manner, we obtained a data set of 15 predicted prices. Figure 3.1 plots actual prices and predicted prices for the study period.

We also regressed the actual prices on the predicted prices using the model:

$$Actual\ price\ =\ \alpha + \beta * Predicted\ price + \epsilon\ . \tag{3.8}$$

The regression results obtained were assessed on the basis of the following criteria proposed by Fama and Gibbons (1984)[]: (1) conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one; (2) serially uncorrelated residuals; and (3) a low residual standard error. Table 3.4 reports the results. On the basis of these results, we can conclude that the intercept is zero [4], although its standard error is quite large, and that the residuals do not display first-order serial correlation. Also, there is a high probability of a slope close to one.[5]

Table 3.4 also reports regression results based on "naive martingale" prediction. In this case, the predicted price for a given month is equated to the actual price for the preceding month. For example, the price prediction for March 1, 1977 is the February 1, 1977 price. What follows from this analysis is the observation that the price predictions emerging from our model are largely

[4]The 95 percent confidence interval for α is (-2.82516, 20.72316).

[5]The 95 percent confidence interval for β is (0.7504804, 0.9775396).

indistinguishable from the "naive martingale" predictions. If we refer back to formula (3.5):

$$E[S_k \mid \mathcal{Y}_{k-1}] = S_{k-1} \sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right) \frac{\langle q_{k-1}, e_i \rangle}{\langle q_{k-1}, \underline{1} \rangle} \ , \ 1 \leq k \leq K \ ,$$

this observation implies that the term:

$$\sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right) \frac{\langle q_{k-1}, e_i \rangle}{\langle q_{k-1}, \underline{1} \rangle} \ -$$

is roughly one. This follows because the values of $g_i(\alpha)$ and $\gamma_i(\alpha)$, $1 \leq i \leq N$, are small; $\exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right)$ is roughly one for $1 \leq i \leq N$; and $\sum_{i=1}^{N} \frac{\langle q_{k-1}, e_i \rangle}{\langle q_{k-1}, \underline{1} \rangle} = 1, 1 \leq k \leq K$. The second column of Table 3.5 reports the values of this term, calculated at the end of each of the fifteen passes through the data.

Further Remarks:

Our model does not appear to capture fully the variability in the data. This is reflected in the earlier observation, reported in Table 3.3, that the variance of the test statistic, our estimate of the noise term in the model, exceeds one. Thus, it appears that the variability in the data is being absorbed to an excessive degree into the noise term rather than into our drift and volatility estimates.

To cast more light on this, we calculated the mean of the $y_k = \log\left(\frac{S_k}{S_{k-1}}\right)$, $1 \leq k \leq 15$, for each pass through the data. We used this as a measure of the drift in prices during each pass. We then plotted these data along with $\hat{g}_{15}$,[6] calculated at the end of each pass, and compared the time paths. See Figure 3.2. This plot does suggest that our estimate of drift understates the true drift in prices.

[6] $\hat{g}_{15}$ is referred to as g.hat in Figure 3.2 .

However, on the positive side, our estimate appears to reflect accurately the direction of price changes; that is, if prices are trending upward our estimate of drift is positive and vice versa.

We also calculated the variance of the y_k, $1 \leq k \leq 15$, for each pass through the data and used this as a measure of price volatility. We then plotted these data along with $\hat{\gamma}_{15}$,[7] calculated at the end of each pass. See Figure 3.3. Again, this suggests that our estimate of volatility understates true volatility but that it accurately reflects the fluctuations in true volatility.

Conclusions:

We can improve the performance of the model by a careful selection of starting values for the model parameters, the time step, and the factor "α" which was used to magnify the price changes. However, our estimates of drift and volatility are too small, although they do appear to reflect accurately the direction of actual drift and volatility changes. As a result, our model does not capture fully the movement in prices and our price predictions are not as successful as we would wish.

3.4.7 Case 2: Gold Prices

In this second case, we analyzed a data set of 90 monthly observations on the price of gold. The sample period ran from November 1988 to April 1996. The data were compiled by Reuters and published in *The Financial Post*.

The data were processed in 15 groups of 6 prices each; at the end of each pass through the data, parameter estimates were updated using the formulas of Chapter 2, section 2.2.1 .

[7] $\hat{\gamma}_{15}$ is referred to as gamma.hat in Figure 3.3 .

Size of N:

In the analysis N, the size of the state space of the Markov chain, was given a value of four. Table 3.6 gives the values of $\hat{g}_K$ and $\hat{\gamma}_K$, as defined in equations (3.6) and (3.7), after the fifteenth pass and for the cases $N = 2$, 4, and 9. Here, K=5 since y_k, $1 \leq k \leq K$, is observed five times in each pass. Note that $\hat{g}_5$ and $\hat{\gamma}_5$ are of the same order of magnitude for the three cases $N = 2$, 4, and 9.

The Time Step:

We again used a time step of 21. Using the alternative time step of 21/252, we observed that $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 5$, which was calculated in each of the 15 passes, assumed the tolerance value of 1.0e-10 278 times out of a possible total of 300. Thus, using the latter time step, information contained in the prices was lost to the parameter estimation.

Starting Values:

The starting values that were assigned to the distribution of the state of the Markov chain, that is, to $E[X_0]$, the transition matrix A, and the vectors g and γ are reported in Table 3.7. We also did the analysis using the starting values based on the continuous time model (discussed in section 3.4.2). Using these latter starting values, we observed that the components of the estimated g and γ vectors were small; for example, the estimated γ vector, after the fifteenth pass, was $(0.001585783, 0.002141250, 0.001776593, 0.022201497)'$. Thus, we rejected these starting values in favour of the former.

Magnifying the Price Change and Checking the Model:

In the analysis, we multiplied the y_k values, $1 \leq k \leq 5$, by a factor of $\alpha = 10.0$ in each pass through the data. We did the analysis using other values of α as well, for example, $\alpha = 1.0$, 2.5, 5.0, 7.5, 20.0, and 100.0 and then selected that value of α that gave the best results based on the considerations enumerated earlier in section 3.4.6. For example, using $\alpha = 10.0$ rather than $\alpha = 1.0$, we were able to reduce the number of times $\Gamma^i(y_k)$ assumed the tolerance value of 1.0e-10 from 173 to 126, out of a possible total of 300. Also, in regard to our two checks on the model, we report, in Table 3.8, the mean and variance of the test statistic:

$$\frac{\log\left(\frac{S_k}{S_{k-1}}\right) - g(\hat{X}_{k-1})}{\gamma(\hat{X}_{k-1})} \ , \ 1 \leq k \leq 5 \ -$$

and the correlation between this statistic and its normal scores, calculated for each pass through the data. These results are reported for the case $\alpha = 10.0$. As we mentioned earlier, this statistic should be normally distributed with a mean of zero and a variance of one. The results of Table 3.8, excluding those for the thirteenth pass, support the hypothesis of normality; as well, they support the hypothesis of zero mean.[8] However, the hypothesis of unit variance is rejected. These results are, however, superior to those obtained for the other values of α. The second check on the model, based on the value of $E[N_k^{r^s} \mid \mathcal{Y}_k]$, is also supported for the case $\alpha = 10.0$.

Parameter Re-estimation:

Table 3.7 gives the re-estimated values of the conditional distribution of the state of the Markov chain, that is, of $E[X_5 \mid \mathcal{Y}_5]$, the transition matrix A, and the

[8]The 95 percent confidence interval for the sample mean is (-12.976, 22.793).

vectors g and γ, after the fifteenth pass through the data. These results are for the case $\alpha = 10.0$.

Price Prediction:

At the end of each pass through the data, we calculated the predicted price for the following month using formula (3.5) and thus obtained a data set of 15 predicted prices. Figure 3.4 plots actual prices and predicted prices for the study period.

As in Case 1, we regressed the actual prices on the predicted prices using model (3.8) and assessed the results obtained using the Fama-Gibbons criteria. Table 3.9 reports the results. They imply that the residuals do not display first-order serial correlation. However, the hypotheses of zero intercept and unit slope are not supported.[9]

Table 3.9 also reports regression results based on "naive martingale" prediction. Here again, as in Case 1, the price predictions arising from our model are largely indistinguishable from the "naive martingale" predictions. This implies that the term:

$$\sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right) \frac{\langle q_{k-1}, e_i \rangle}{\langle q_{k-1}, \underline{1} \rangle} \ , \ 1 \leq k \leq 5 -$$

that occurs in formula (3.5) is roughly one. The third column of Table 3.5 reports the values of this term, calculated at the end of each of the fifteen passes through the data.

[9]The 95 percent confidence intervals for α and β, respectively, are (20.4036, 175.0164) and (0.5219136, 0.9378864).

Further Remarks:

As in Case 1, it appears that our model does not capture fully the variability in the data. Note that the gold price data are more variable than the IBM stock price data.

We again calculated the mean of $y_k = \log\left(\frac{S_k}{S_{k-1}}\right), 1 \leq k \leq 5$, for each pass through the data and used this as a measure of the drift in prices. We plotted these data along with $\hat{g}_5$[10] and compared their time paths. See Figure 3.5. This plot suggests that our estimate of drift understates the true drift in prices but that it accurately reflects the direction of price trends.

We also calculated the variance of y_k, $1 \leq k \leq 5$, for each pass through the data and, using this as a measure of price volatility, plotted these data along with $\hat{\gamma}_5$ in Figure 3.6.[11] This plot suggests that our estimate of volatility understates true price volatility but that it accurately reflects the fluctuations in true volatility.

Conclusions:

We conclude that our estimates of drift and volatility are too small, although they do appear to accurately reflect the direction of actual drift and volatility changes.

3.4.8 Case 3: United States Dollar Prices

In this third case, we analyzed a data set of 90 monthly observations on the price of the United States (U.S.) dollar relative to the Canadian dollar. The sample

[10] $\hat{g}_5$ is referred to as g.hat in Figure 3.5.

[11] $\hat{\gamma}_5$ is referred to as gamma.hat in Figure 3.6.

period ran from October 1988 to March 1996. The data were compiled by the Royal Bank of Canada and published in *The Financial Post.*

The data were processed in 15 groups of 6 prices each; at the end of each pass through the data, parameter estimates were updated using the formulas of Chapter 2, section 2.2.1.

Size of N:

In the analysis, we let N take the value four. Table 3.10 gives the values of $\hat{g}_5$ and $\hat{\gamma}_5$, as defined in equations (3.6) and (3.7), after the fifteenth pass and for the cases $N = 2$, 4, and 9.

The Time Step:

We again used a time step of 21. Using the alternative time step of 21/252, we observed that $\Gamma^i(y_k), 1 \leq i \leq 4, 1 \leq k \leq 5$, which was calculated in each of the 15 passes, assumed the tolerance value of 1.0e-10 278 times out of a possible total of 300.

Starting Values:

See Table 3.11 for the starting values that were assigned to the distribution of the state of the Markov chain, that is, to $E[X_0]$, the transition matrix A, and the vectors g and γ. We also did the analysis using the starting values based on the continuous time model (discussed in section 3.4.2). Using these latter starting values, we observed that the components of the estimated g and γ vectors were small; for example, the estimated γ vector,after the fifteenth pass, was $(0.0001226494, 0.0003273091, 0.0001618913, 0.0001520375)'$.

Magnifying the Price Changes and Checking the Model:

In the analysis, we multiplied the y_k values, $1 \leq k \leq 5$, by a factor of $\alpha = 5.0$ in each pass through the data. We did the analysis using other values of α as well and then selected that value of α that gave the best results based on the considerations given earlier in section 3.4.6. For example, using $\alpha = 5.0$ rather than $\alpha = 1.0$, we were able to reduce the number of times $\Gamma^i(y_k)$ assumed the tolerance value of 1.0e-10 from 224 to 181, out of a possible total of 300. Also, in regard to the two checks on the model, we report, in Table 3.12, the mean and variance of the test statistic:

$$\frac{\log\left(\frac{S_k}{S_{k-1}}\right) - g(\hat{X}_{k-1})}{\gamma(\hat{X}_{k-1})} \ , 1 \leq k \leq 5 \ -$$

and the correlation between this statistic and its normal scores, calculated for each pass through the data. These results are reported for the case $\alpha = 5.0$. Except for the third, ninth, and twelth passes, the results of Table 3.12 support the hypothesis that the statistic is normally distributed. Also, the statistic has zero mean.[12] However, the hypothesis of unit variance is not supported. These results are superior to those obtained for other values of α. The second check on the model, based on the value of $E[N_k^{r^s} \mid \mathcal{Y}_k]$, is also supported for the case $\alpha = 5.0$.

Parameter Re-estimation:

Table 3.11 gives the re-estimated values of the conditional distribution of the state of the Markov chain, that is, of $E[X_5 \mid \mathcal{Y}_5]$, the transition matrix A, and the vectors g and γ, after the fifteenth pass through the data. These results are for the case $\alpha = 5.0$.

[12]The 95 percent confidence interval for the sample mean is (-12.9756, 22.7934).

Price Prediction:

At the end of each pass through the data, we calculated the predicted price for the following month using formula (3.5) and thus obtained a data set of 15 predicted prices. Figure 3.7 plots actual and predicted prices for the study period.

As in Cases 1 and 2, we regressed the actual prices on the predicted prices using model (3.8) and assessed the results obtained using the Fama-Gibbons criteria. Table 3.13 reports the results. They imply that the intercept is zero, [13] the slope is one,[14] and the residuals are serially uncorrelated.

Table 3.9 also reports regression results based on "naive martingale" prediction. As in Cases 1 and 2, the price predictions arising from our model are largely indistinguishable from the "naive martingale" predictions. The fourth column of Table 3.5 gives the values of the term:

$$\sum_{i=1}^{N} \exp\left(\frac{g_i(\alpha)}{\alpha} + \frac{\gamma_i^2(\alpha)}{2\alpha^2}\right) \frac{\langle q_{k-1}, e_i\rangle}{\langle q_{k-1}, \underline{1}\rangle}\ ,\ 1 \leq k \leq 5 -$$

that occurs in formula (3.5) for $E[S_k \mid \mathcal{Y}_{k-1}]$. This term is reported for each of the fifteen passes through the data. As can be observed, it takes values close to one.

Further Remarks:

Figures 3.8 and 3.9 support our earlier observations, arising from the analyses of IBM stock and gold prices, that our model does not capture fully the variability in the data, although the data here are less variable than in the previous two

[13]The 95 percent confidence interval for α is (-0.1262092, 0.1702292).

[14]The 95 percent confidence interval for β is (0.867156, 1.103244).

cases. Estimates of price drift and price volatility appear to understate the true drift and volatility in prices. However, the direction of drift and volatility changes appear to be captured accurately in our estimates.

Conclusions:

Our conclusions are the same as those of Cases 1 and 2.

TABLE 3.1: $\hat{g}_{15}$ and $\hat{\gamma}_{15}$, after the fifteenth pass, for selected values of N – IBM stock prices

	N=2	N=4	N=9
$\hat{g}_{15}$	0.025564643	0.037539863	0.022820964
$\hat{\gamma}_{15}$	0.006360731	0.007960336	0.005953758

TABLE 3.2: **Starting values and parameter estimates – IBM stock prices**

Starting values:

$E[X_0]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:
$$\begin{pmatrix} 0.23 & 0.27 & 0.20 & 0.30 \\ 0.20 & 0.30 & 0.30 & 0.27 \\ 0.30 & 0.20 & 0.27 & 0.20 \\ 0.27 & 0.23 & 0.23 & 0.23 \end{pmatrix}$$

g vector: (0.9, 0.1, 0.8, 0.2)

γ vector: (0.25, 0.25, 0.25, 0.25)

After the fifteenth pass:

$E[X_{15} \mid \mathcal{Y}_{15}]$ vector: (0.55317729, 0.35354643, 0.00091133, 0.09236493)

A matrix:
$$\begin{pmatrix} 1.094e-11 & 0.5303 & 3.372e-07 & 8.820e-08 \\ 9.995e-01 & 0.4527 & 9.999e-01 & 9.999e-01 \\ 4.334e-04 & 0.0049 & 2.164e-08 & 6.093e-10 \\ 8.976e-06 & 0.0119 & 3.408e-06 & 2.028e-11 \end{pmatrix}$$

g vector: (0.05846258, 0.02057468, -0.05771766, -0.02188939)

γ vector: (0.00566290, 0.01352559, 0.00310218, 0.00046546)

TABLE 3.3: **The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – IBM stock prices**

Pass number	Mean	Variance	Correlation
1	4.0679480	443.8428	0.9671332
2	-1.4273101	455.9257	0.9590864
3	0.4883605	530.3145	0.9616368
4	0.8699806	294.5432	9.9644653
5	-4.2289792	907.3855	0.9659001
6	0.3639345	238.4010	0.9849245
7	0.5278630	289.6081	0.9118571
8	6.3914976	1649.2339	0.9724559
9	0.5370361	4423.4478	0.9645612
10	1.9824365	204.5500	0.9670112
11	-3.4370434	676.7271	0.9617791
12	0.6918717	401.0032	0.9889625
13	-2.6111324	106.5434	0.9302095
14	1.0123313	282.5690	0.9734133
15	4.1499271	381.6925	0.9826179

TABLE 3.4: **Regressions of actual prices on predicted prices – IBM stock prices**

Parameter	The model	Martingale prediction
α	8.949 (5.451)	8.610 (5.582)
β	0.86401 (0.05256)	0.86921 (0.05396)
R-squared	0.954	0.952
Durbin-Watson D statistic	2.25	2.13
s	7.254	7.395

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 3.5: **The term that is multiplied by S_{k-1} to get $E[S_k \mid \mathcal{Y}_{k-1}]$, calculated at the end of each pass through the data**

Pass number	IBM stock prices	Gold prices	U.S. dollar prices
1	1.006418	0.998674	0.999401
2	0.999028	0.999419	0.999549
3	1.002490	0.999440	1.000052
4	0.997722	1 000562	0.999479
5	1.002435	0.998346	0.999830
6	1.021541	1.000247	0.999404
7	1.003608	0.997265	1.001900
8	0.999131	1.002309	1.000132
9	1.001876	0.998845	1.000118
10	0.994414	1.000939	1.001943
11	1.005082	1.001163	1.000738
12	0.989428	1.000565	0.999563
13	0.975443	1.000213	1.001373
14	1.001213	0.999272	0.998569
15	1.015191	1.000814	1.001388

TABLE 3.6: $\hat{g}_5$ and $\hat{\gamma}_5$, after the fifteenth pass, for selected values of N – gold prices

	N=2	N=4	N=9
$\hat{g}_5$	0.0065892649	0.008115679	0.0071034259
$\hat{\gamma}_5$	0.0144979179	0.0161086519	0.0170995747

TABLE 3.7: **Starting values and parameter estimates – gold prices**

Starting values:

$E[X_0]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:
$$\begin{pmatrix} 0.23 & 0.27 & 0.20 & 0.30 \\ 0.20 & 0.30 & 0.30 & 0.27 \\ 0.30 & 0.20 & 0.27 & 0.20 \\ 0.27 & 0.23 & 0.23 & 0.23 \end{pmatrix}$$

g vector: (0.9, 0.1, 0.8, 0.2)

γ vector: (0.25, 0.25, 0.25, 0.25)

After the fifteenth pass:

$E[X_5 \mid \mathcal{Y}_5]$ vector: (0.8363017, 0.1633632, 0.0001140, 0.0002209)

A matrix:
$$\begin{pmatrix} 5.864\text{e}{-01} & 9.992\text{e}{-01} & 9.695\text{e}{-02} & 9.999\text{e}{-01} \\ 4.133\text{e}{-01} & 2.845\text{e}{-06} & 9.029\text{e}{-01} & 3.099\text{e}{-09} \\ 2.793\text{e}{-07} & 7.284\text{e}{-04} & 2.971\text{e}{-15} & 1.324\text{e}{-07} \\ 1.819\text{e}{-04} & 6.330\text{e}{-05} & 8.356\text{e}{-05} & 2.145\text{e}{-15} \end{pmatrix}$$

g vector: (0.01146567, -0.00900794, 0.01776684, -0.01602404)

γ vector: (0.01879905, 0.00234592, 0.00538968, 0.01411942)

TABLE 3.8: **The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – gold prices**

Pass number	Mean	Variance	Correlation
1	-18.6914141	1489.4856	0.9793188
2	-5.9482811	1242.3133	0.9761170
3	-0.5587373	147.8234	0.9838245
4	19.7914630	2277.4604	0.9841864
5	-8.4694556	783.1129	0.9556415
6	-14.0825716	34990.0607	0.9208052
7	-6.6966485	849.2314	0.9842449
8	5.9121606	806.3495	0.9587379
9	-52.3198635	35832.8194	0.9736994
10	-2.8389682	401.4180	0.9635492
11	33.5893993	4903.8583	0.9680359
12	87.1144038	23598.5014	0.9545966
13	46.7625758	14303.3264	0.7718421
14	-13.2721931	10657.8872	0.9737155
15	3.3429573	710.5670	0.9533481

TABLE 3.9: **Regressions of actual prices on predicted prices – gold prices**

Parameter	The model	Martingale prediction
α	98.00 (36.41)	95.59 (36.42)
β	0.72880 (0.09794)	0.73522 (0.09795)
R-squared	0.810	0.813
Durbin-Watson D statistic	1.75	1.74
s	7.589	7.536

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 3.10: $\hat{g}_5$ and $\hat{\gamma}_5$, after the fifteenth pass, for selected values of N – U.S. dollar prices

	N=2	N=4	N=9
$\hat{g}_5$	0.0053162479	0.0069351150	0.0060696187
$\hat{\gamma}_5$	0.0002449577	0.0002750454	0.0002549512

TABLE 3.11: **Starting values and parameter estimates – U.S. dollar prices**

Starting values:

$E[X_0]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:
$$\begin{pmatrix} 0.23 & 0.27 & 0.20 & 0.30 \\ 0.20 & 0.30 & 0.30 & 0.27 \\ 0.30 & 0.20 & 0.27 & 0.20 \\ 0.27 & 0.23 & 0.23 & 0.23 \end{pmatrix}$$

g vector: (0.9, 0.1, 0.8, 0.2)

γ vector: (0.25, 0.25, 0.25, 0.25)

After the fifteenth pass:

$E[X_5 \mid \mathcal{Y}_5]$ vector: (0.04875742, 0.73922200, 0.10572331, 0.10629726)

A matrix:
$$\begin{pmatrix} 0.3403128 & 0.4608361 & 0.2322707 & 0.33886943 \\ 0.1195731 & 0.1484816 & 0.4857810 & 0.46581355 \\ 0.2981371 & 0.1745457 & 0.1555276 & 0.09744800 \\ 0.2419771 & 0.2161366 & 0.1264207 & 0.09786903 \end{pmatrix}$$

g vector: (0.00560040, 0.00863796, 0.00140784, 0.00120267)

γ vector: (0.00018385, 0.00032621, 0.00012265, 0.00011262)

TABLE 3.12: **The mean and variance of the test statistic and its correlation with its normal scores, calculated for each pass through the data – U.S. dollar prices**

Pass number	Mean	Variance	Correlation
1	-13.8703781	6913.785	0.9548172
2	-16.2439767	10332.270	0.9311556
3	0.3138419	7730.959	0.8176349
4	-32.9241847	20968.686	0.9646766
5	-7.7404597	21351.889	0.9956543
6	-60.5868133	26576.208	0.9287732
7	52.1525553	5402.729	0.9342805
8	7.3336312	16832.732	0.9688997
9	1.2685814	3754.749	0.8297596
10	51.3008225	3642.245	0.9508829
11	9.7377676	6056.353	0.9765218
12	-38.4199756	24104.886	0.8676565
13	31.8379451	5985.651	0.9819447
14	-29.9925761	5348.683	0.9916077
15	55.0524817	12091.701	0.9889897

TABLE 3.13: **Regressions of actual prices on predicted prices – U.S. dollar prices**

Parameter	The model	Martingale prediction
α	0.02201 (0.06862)	0.01707 (0.06997)
β	0.98520 (0.05465)	0.98940 (0.05575)
R-squared	0.962	0.960
Durbin-Watson D statistic	1.98	1.94
s	0.01920	0.01949

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

FIGURE 3.1 Actual and predicted prices of IBM stock

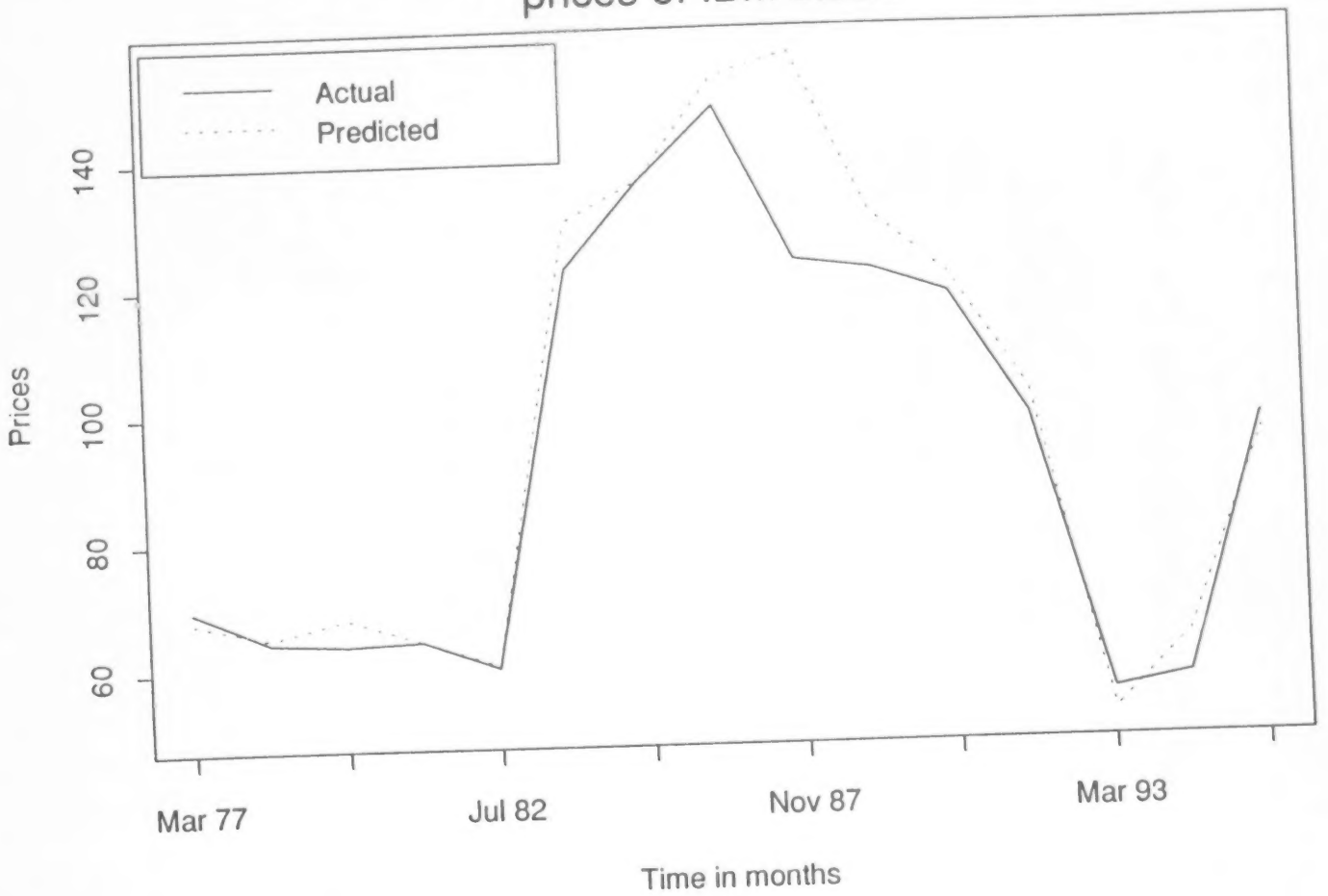


FIGURE 3.2 Mean of y versus g.hat - IBM stock prices

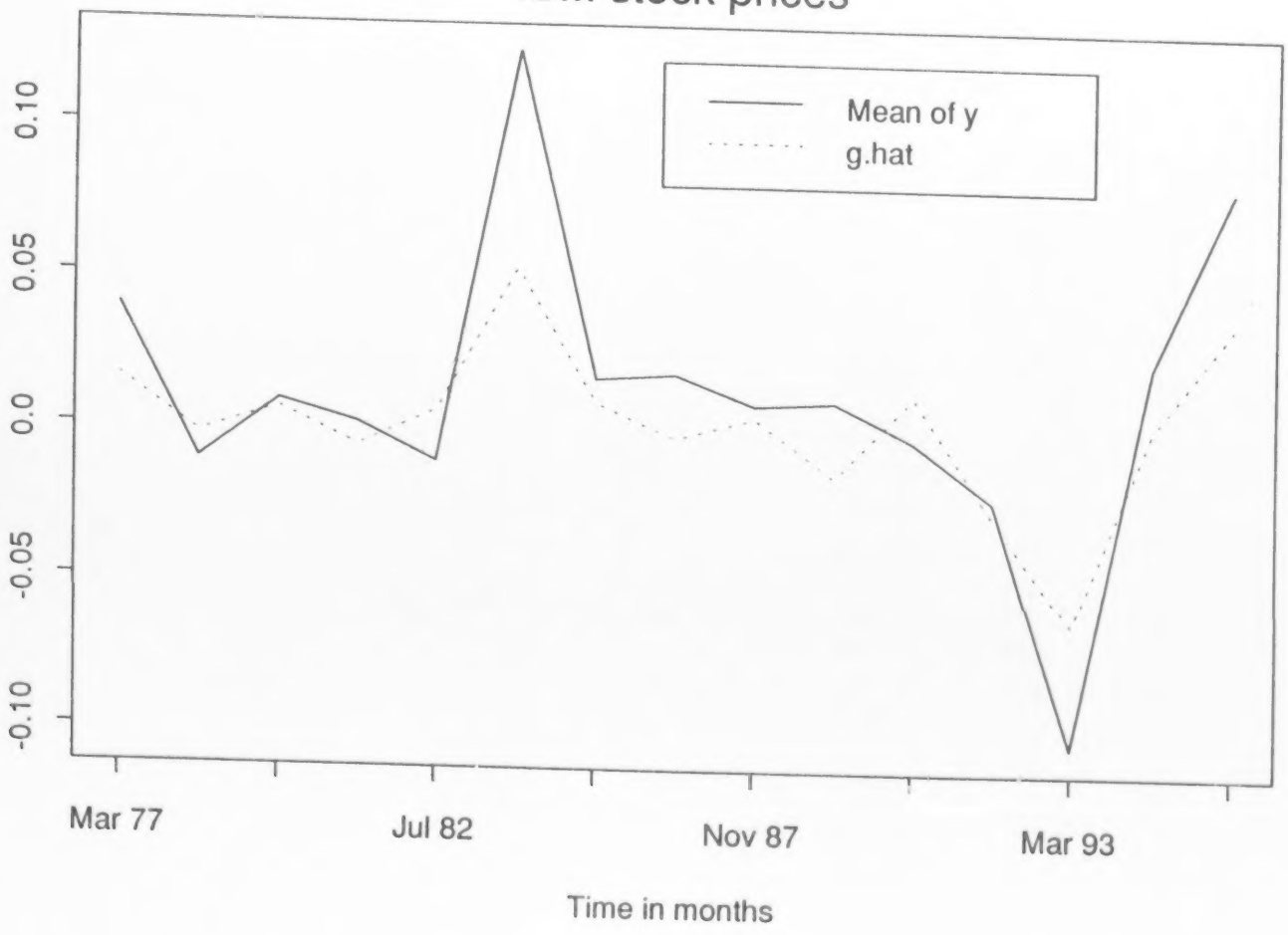


FIGURE 3.3 Variance of y versus gamma.hat - IBM stock prices

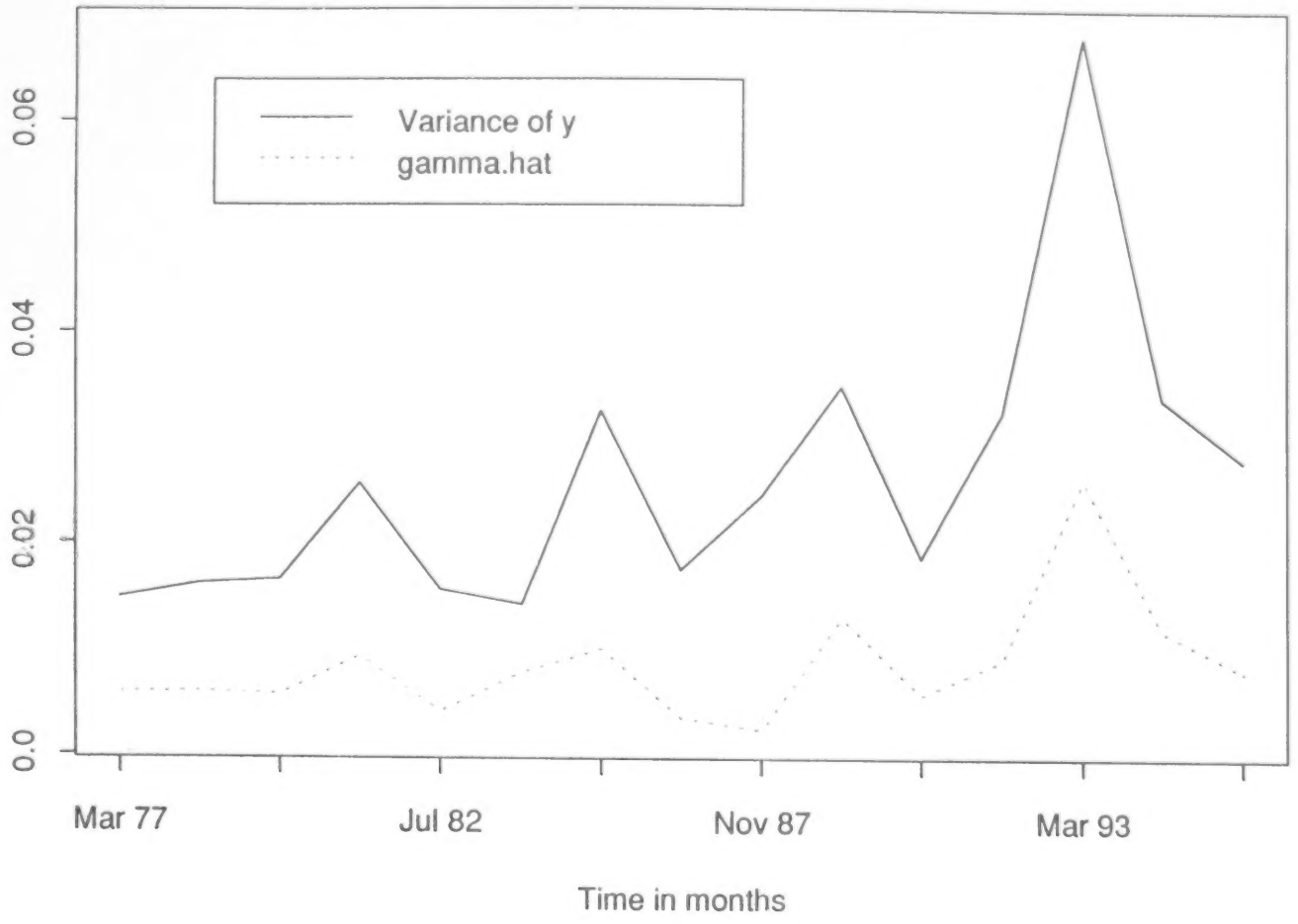


FIGURE 3.4 Actual and predicted prices of gold

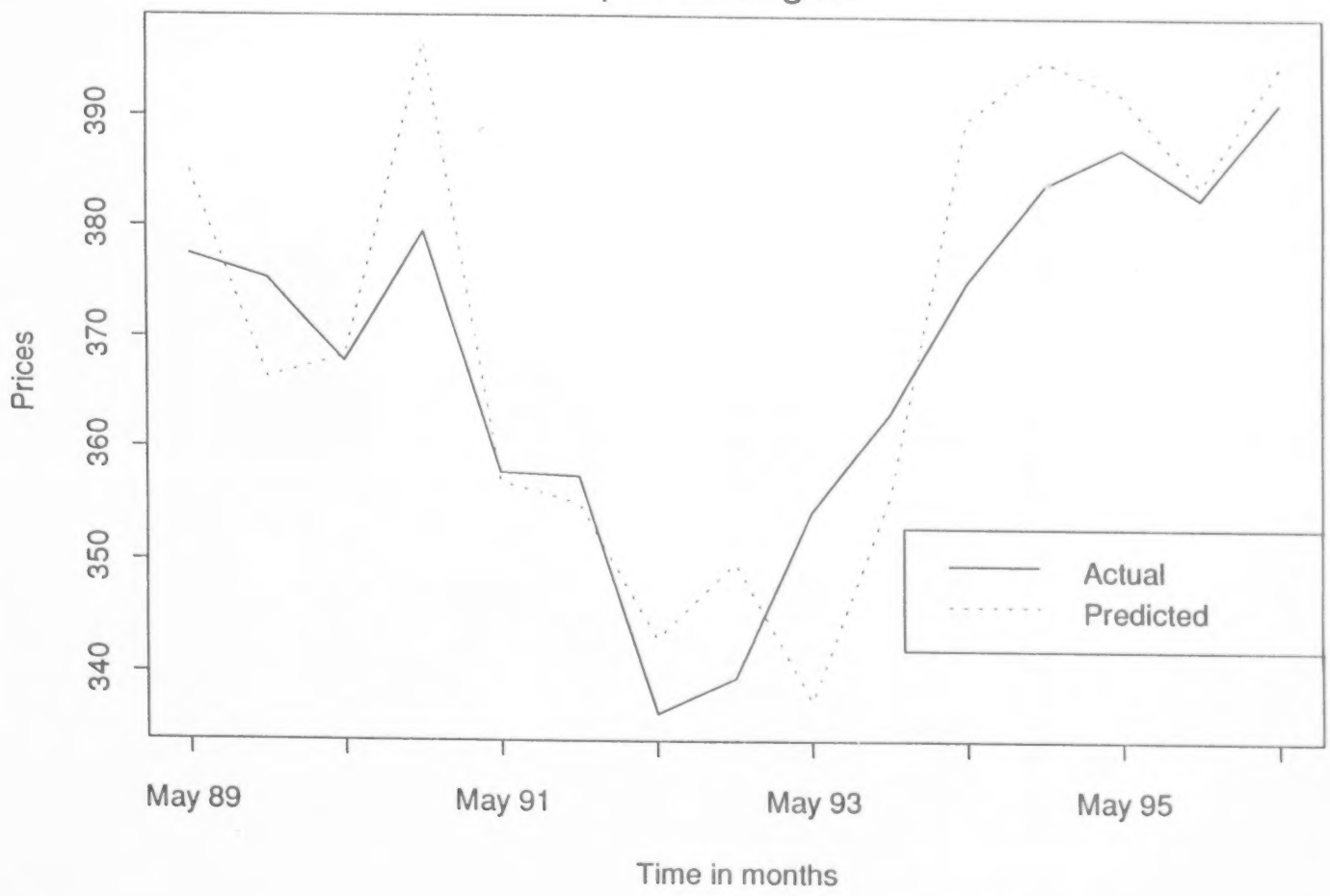


FIGURE 3.5 Mean of y versus g.hat - gold prices

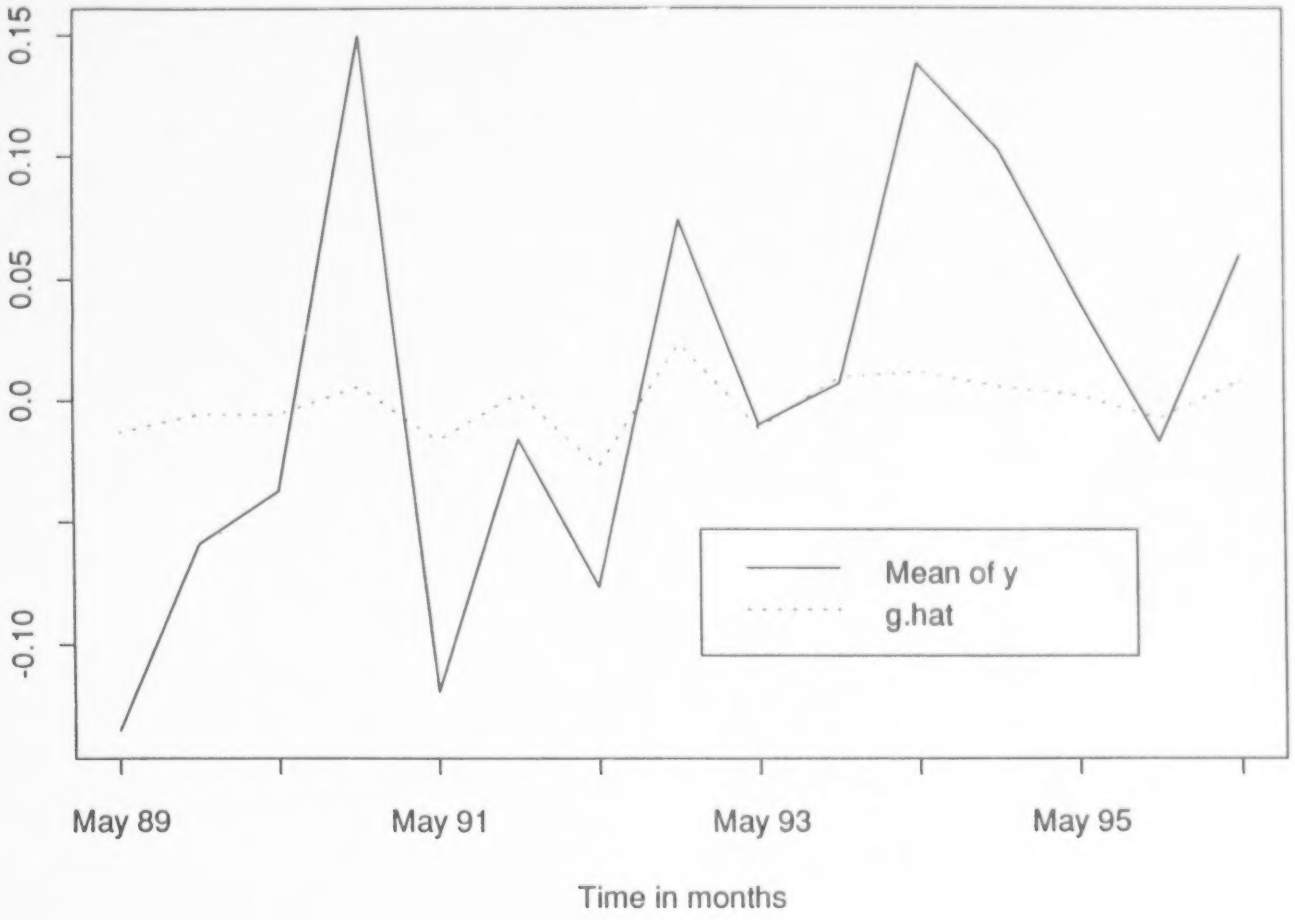


FIGURE 3.6 Variance of y versus gamma.hat - gold prices

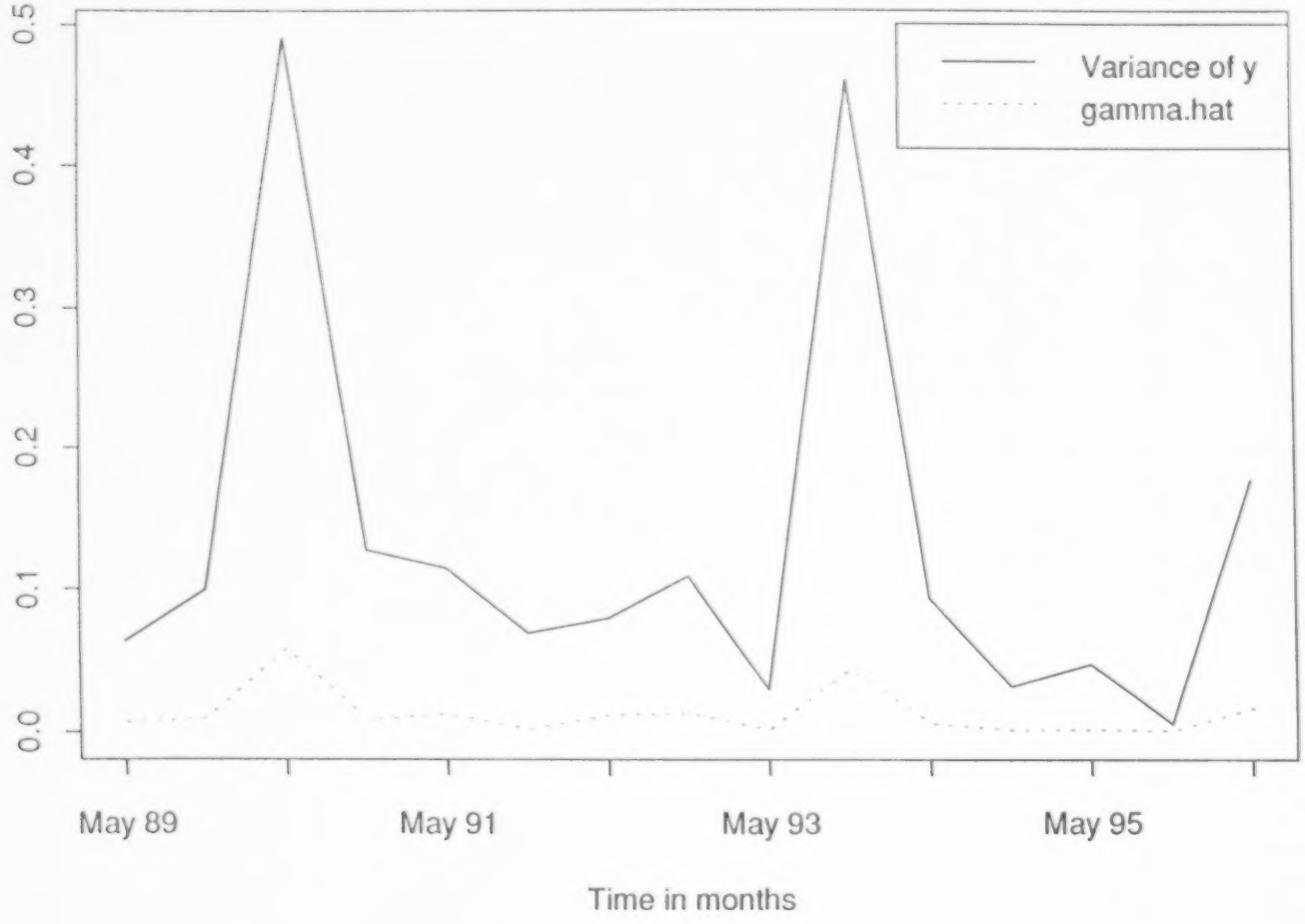


FIGURE 3.7 Actual and predicted prices of U.S. dollar

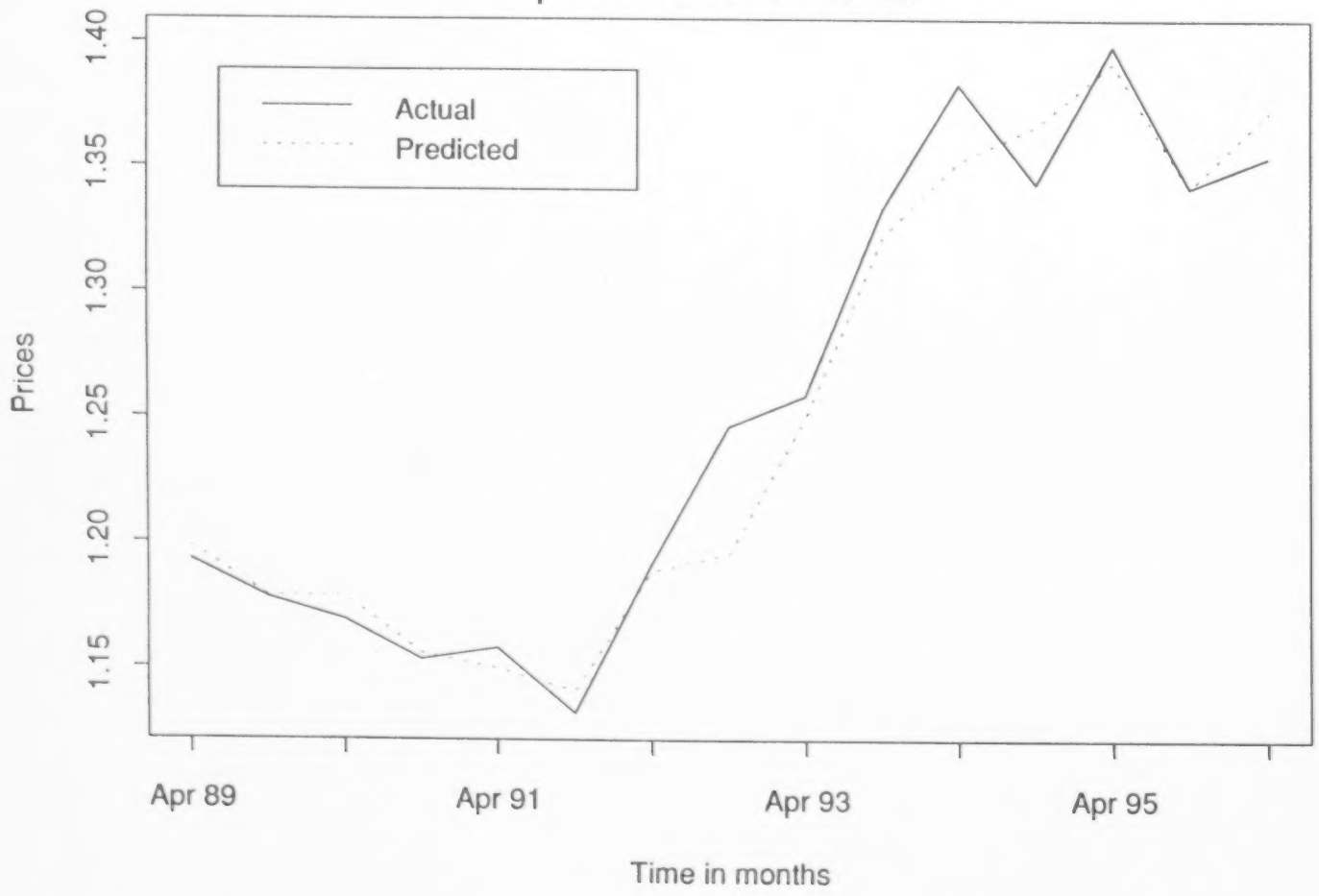


FIGURE 3.8 Mean of y versus g.hat -
U.S. dollar prices

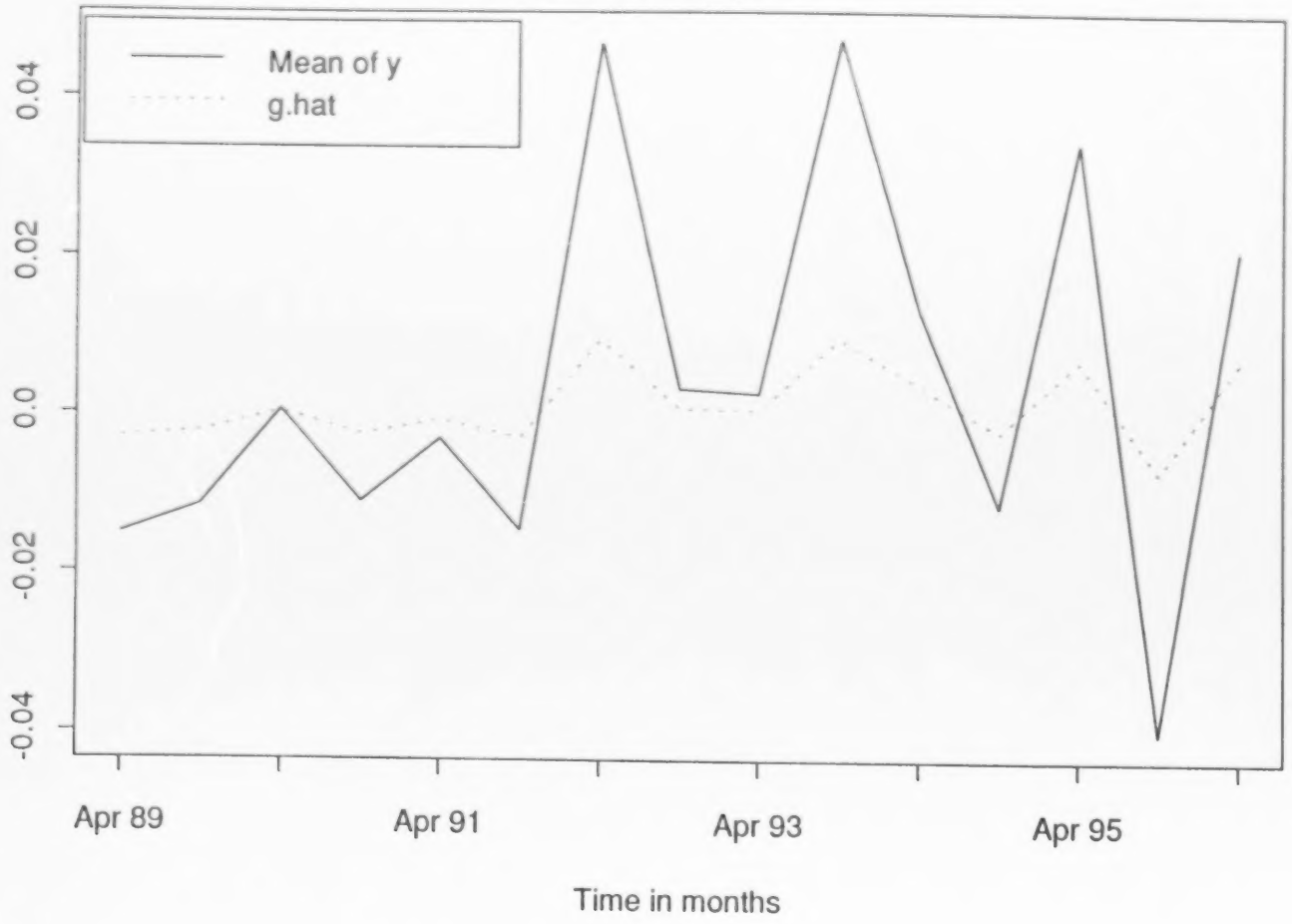
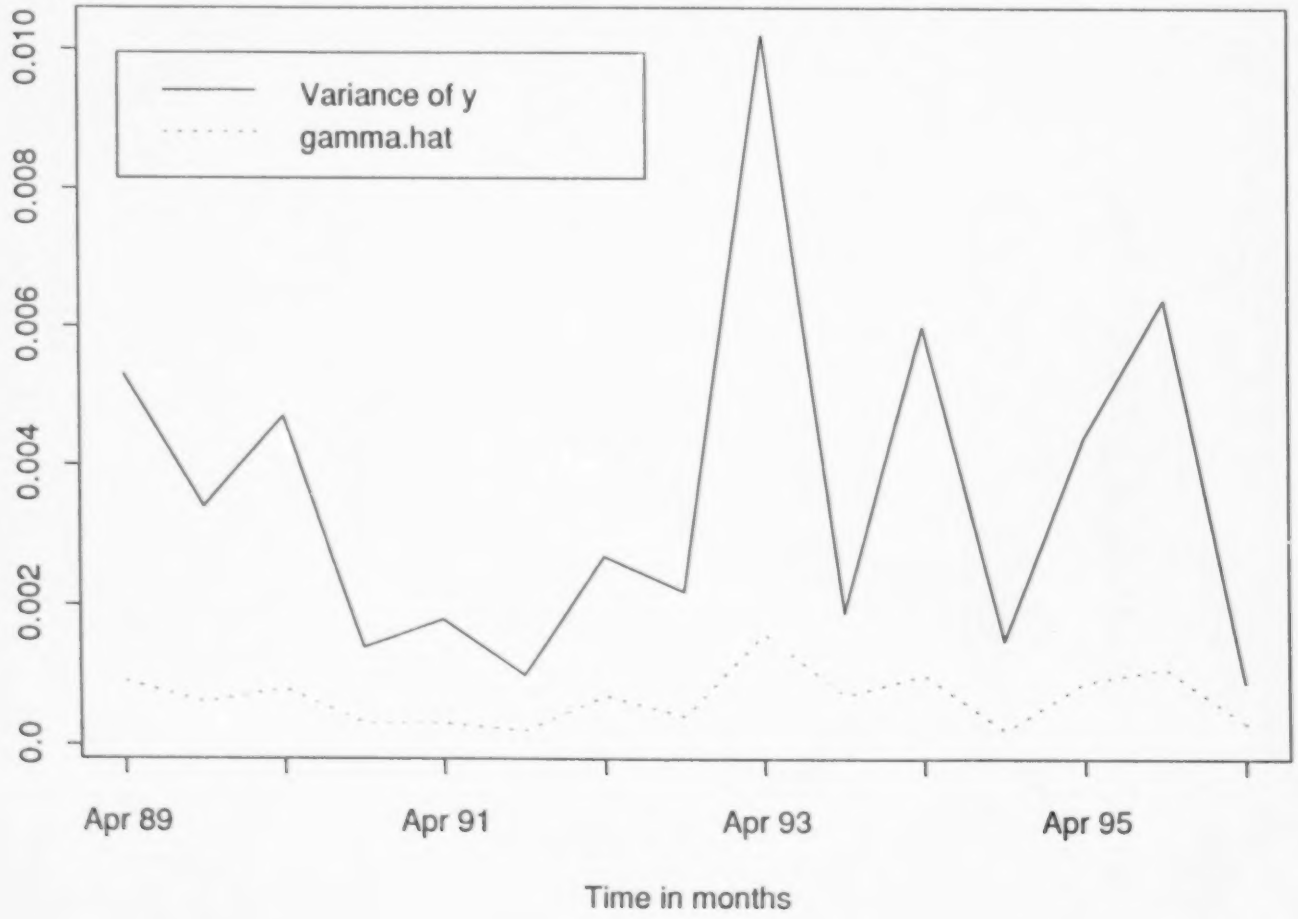


FIGURE 3.9 Variance of y versus gamma.hat - U.S. dollar prices



Chapter 4

Application: Modelling the Term Structure of Interest Rates

This chapter applies the theory of Chapter 2 to the problem of modelling the term structure of interest rates.[1] It makes some preliminary comments and then discusses the model, an application of the model, and results.

4.1 Preliminaries

In this chapter, we assume that all processes are defined on a complete probability space $(\Omega, \mathcal{F}, P)$. We first recall the following:

Definition 4.1.1 Two probabilities P *and* Q are equivalent *if and only if, for all events* $A \subset \Omega$, $P(A) = 0$ *if and only if* $Q(A) = 0$.

[1] Important references in this chapter are Elliott, Aggoun, and Moore [8] , Elliott [7] , Elliott [6], and Elliott, Hunter, and Jamieson [11] . The discussion of the first two sections follows Elliott, Hunter, and Jamieson [11].

We also assume that there exists a probability measure Q, equivalent to P, such that under Q the discounted prices of all securities, Treasury bills and bonds in this case, are martingales.

4.2 The Model

4.2.1 Instantaneous Rate Process

Having assumed the existence of an equivalent martingale measure Q, we suppose that all processes are defined on probability space $(\Omega, \mathcal{F}, Q)$. On this space, let $X = \{X_t; t \geq 0\}$ be a Markov chain taking on values in a finite set S. Without loss of generality, we can identify the points in S with the unit vectors in $\mathbb{R}^N$, where N is the cardinality of S. Let $\Sigma = \{e_1, \ldots e_N\}$ denote the set of standard unit vectors in $\mathbb{R}^N$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$. Because $X_t \in \Sigma$ for $t \geq 0$, any function of X_t, say $h(X_t)$, is given by a vector $h = (h_1, \ldots, h_N)'$ and $h(X_t) = \langle h, X_t \rangle$. Letting E represent expectation under Q, the distribution of X_t is then $E[X_t] = p_t = (p_t^1, \ldots, p_t^N)'$, where $p_t^i = Q(X_t = e_i) = E[\langle X_t, e_i \rangle]$. We suppose that this distribution evolves according to the Kolmogorov equation:

$$\frac{dp_t}{dt} = Cp_t$$

where $C = (c_{ji})$, $1 \leq i, j \leq N$, $\sum_{j=1}^{N} c_{ji} = 0$, and $c_{ji} \geq 0$ if $i \neq j$.

Furthermore, we suppose that the instantaneous rate process, $r = \{r_t\}$,[2] has the property that r_t is a function of X_t, $t \geq 0$; that is, $r_t = r(X_t) = \langle r, X_t \rangle$ for some vector $r \in \mathbb{R}^N$.

We let $\{\mathcal{F}_t\}$, $t \geq 0$, represent the right continuous, complete filtration generated by X. In chapters 1 and 2, we defined the terms generated σ-field and

[2]See Chapter 1, section 1.3.1 for a definition of the *instantaneous interest rate*.

completeness, as well as the term filtration, for the discrete-time case. The definition of filtration, for the continuous-time case, is similar.

Definition 4.2.1 *On* $(\Omega, \mathcal{F}, P)$, *a* filtration *is an increasing family of sub* σ-*fields* $\mathcal{F}_t \subset \mathcal{F}$. *Increasing means that if* $s \leq t$, *then* $\mathcal{F}_s \subset \mathcal{F}_t$, $s, t \geq 0$.

Right continuous *means that* $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

If we consider a security paying a non-random amount u at time $T \geq t$, then its price at time t is:

$$E\left[\exp\left(-\int_t^T r(X_s)ds\right) u \mid \mathcal{F}_t\right] = E\left[\exp\left(-\int_t^T r(X_s)ds\right) u \mid X_t\right] ,$$

since X is a Markov process. Taking $u = 1$, the price of a zero-coupon bond maturing at time T is:

$$p_{t,T} = E\left[\exp\left(-\int_t^T r(X_s)ds\right) \mid X_t\right].$$

The yield for this bond is:

$$y_{t,T} = \frac{-1}{(T-t)} \log p_{t,T} .$$

4.2.2 Bond Dynamics

Suppose a security pays a random amount $u(X_T)$ at time $T \geq t$. Then its price at time t, denoted by $F(X_t, t)$, is:

$$F(X_t, t) \;=\; E\left[\exp\left(-\int_t^T r(X_s)ds\right) u(X_T) \mid X_t\right] .$$

This is a function of X_t. Therefore, there exists a function $\psi_t = (\psi_t^1, \ldots, \psi_t^N)' \in \mathbb{R}^N$, with $\psi_t^i = F(e_i, t)$, such that:

$$F(X_t, t) = \langle \psi_t, X_t \rangle .$$

Definition 4.2.2 *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space with filtration* $\{\mathcal{F}_t\}$, $t \geq 0$. *An adapted stochastic process* $M = \{M_t\}$ *is said to be a* martingale (*or a* martingale relative to the filtration $\{\mathcal{F}_t\}$) *if* $E[|M_t|] < \infty$ *for all* t *and* $E[M_t \mid \mathcal{F}_s] = M_s, s \leq t$.

Lemma 4.2.3 *The discounted price of the above-mentioned security is an* $\mathcal{F}_t$*-martingale under* Q.

Proof: Let E denote expectation under Q. Then:

$$\begin{aligned} & \exp\left(-\int_0^t r(X_s)ds\right) F(X_t, t) \\ = \; & \exp\left(-\int_0^t r(X_s)ds\right) E\left[\exp\left(-\int_t^T r(X_s)ds\right) u(X_T) \mid X_t\right] \\ = \; & E\left[\exp\left(-\int_0^T r(X_s)ds\right) u(X_T) \mid \mathcal{F}_t\right] . \end{aligned}$$

■

Lemma 4.2.4 $M_t := X_t - X_0 - \int_0^t CX_u du$ *is an* $\mathcal{F}_t$*-martingale under* Q.

Proof: To start, we write $\Phi(t,s) = \exp C(t-s)$ for the transition matrix associated with C.[3] Therefore, we have:

$$\frac{d}{dt}\Phi(t,s) = C\Phi(t,s) \tag{4.1}$$

and for $s \leq t$:

$$\begin{aligned} E[X_t \mid \mathcal{F}_s] &= E[X_t \mid X_s] \\ &= \Phi(t,s)X_s . \end{aligned}$$

[3]Note that e^M represents $\sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) M^n$ whenever M is a square matrix.

Now:

$$\begin{aligned} & E[M_t - M_s \mid \mathcal{F}_s] \\ = \; & E\left[X_t - X_s - \int_s^t CX_u du \mid X_s\right] \\ = \; & \Phi(t,s)X_s - X_s - \int_s^t C\Phi(u,s)X_s du \\ = \; & \underline{0} \in \mathbb{R}^N, \text{ by (4.1)} . \end{aligned}$$

■

Remark 4.2.5 The semimartingale representation of the Markov chain X is, therefore:

$$X_t = X_0 + \int_0^t CX_u du + M_t .$$

□

Remark 4.2.6 We observe the following:

1. Using the representations:

   - $F(X_s, s) = \langle \psi_s, X_s \rangle$, and
   - $X_t = X_0 + \int_0^t CX_u du + M_t$

   and a general stochastic differentiation rule, it follows that:

$$\begin{aligned} & \exp\left(-\int_0^t r(X_s)ds\right) F(X_t, t) = F(X_0, 0) \\ + \; & \int_0^t -r(X_s)\exp\left(-\int_0^s r(X_u)du\right)\langle \psi_s, X_s\rangle ds \\ + \; & \int_0^t \exp\left(-\int_0^s r(X_u)du\right)\left[\langle \frac{d\psi_s}{ds}, X_s\rangle + \langle \psi_s, CX_s\rangle\right] ds \\ + \; & \int_0^t \exp\left(-\int_0^s r(X_u)du\right)\langle \psi_s, dM_s\rangle . \end{aligned} \tag{4.2}$$

2. Here we have the representation $F(X_s, s) = \langle \psi_s, X_s \rangle$ so that $dF(X_s, s) = \langle d\psi_s, X_s \rangle + \langle \psi_s, dX_s \rangle$.

3. Since the left side of (4.2) is a martingale, it follows that the bounded variation terms on the right side, that is, the ds integrals, give the identically zero process; that is:

$$\exp\left(-\int_0^s r(X_u)du\right)\left[-r(X_s)\langle\psi_s, X_s\rangle + \langle\frac{d\psi_s}{ds}, X_s\rangle + \langle\psi_s, CX_s\rangle\right] = 0 \quad (4.3)$$

4. $r(X_s) = \langle r, X_s \rangle$, where $r = (r_1, \ \ldots \ , r_N)'$, and $r(X_s)\langle\psi_s, X_s\rangle = \langle diag\ r \cdot \psi_s, X_s\rangle$, where $diag\ r$ is the matrix with r on the diagonal. Therefore, from (4.3), we have:

$$\langle\frac{d\psi_s}{ds}, X_s\rangle + \langle C^*\psi_s, X_s\rangle - \langle diag\ r \cdot \psi_s, X_s\rangle = 0 \qquad (4.4)$$

for all X_s.

5. From (4.4), it follows that ψ_s is given by the system of equations:

$$\frac{d\psi_t}{dt} = (diag\ r - C^*)\psi_t \qquad (4.5)$$

with terminal condition:

$$\psi_T = u = (u_1, \ \ldots \ , u_N)' \qquad (4.6)$$

where $u(X_T) = \langle u, X_T \rangle$.

□

Letting $B = diag\ r - C^*$ so $\psi_t = \exp^{-B(T-t)}u$, the price at time $t \leq T$ of a security paying $u(X_T)$ at time T is:

$$\langle \psi_t, X_t \rangle = \langle e^{-B(T-t)}u, X_t \rangle \ .$$

Furthermore, a zero-coupon bond, with $u = \underline{1} \in \mathbb{R}^N$, has price at time $t \leq T$ of:

$$F(X_t, t) = \langle e^{-B(T-t)}\underline{1}, X_t \rangle \ .$$

4.2.3 Filtering and Model Estimation

To enable discussion of the term structure of interest rates, we now introduce bonds of varying maturities. Let $F^j(X_t, t)$ denote the price at time t of a zero-coupon bond expiring at time $t + \tau_j$, $1 \leq j \leq m$. Then, by the analysis of the previous section:

$$\begin{aligned} F^j(X_t, t) &= E\left[\exp\left(-\int_t^{t+\tau_j} r(X_s)ds\right) \mid X_t\right] \\ &= \langle e^{-B\tau_j}\underline{1}, X_t \rangle \ , \ 1 \leq j \leq m \ . \end{aligned}$$

In this model, the process X is not observed directly. We suppose that X represents the dynamics of some factors which are not observed and that the $F^j(X_t, t)$ represent some corresponding 'perfect' prices. In fact, we suppose that the prices $F^j(X_t, t)$, $1 \leq j \leq m$, are observed in noise at discrete times $t_1, t_2, \ \ldots \ , t_k, \ \ldots \ .$ Since $F^j(X_t, t) > 0$, we assume that the noise is multiplicative; that is, we assume that we actually observe $F^j(X_t, t)e^{\langle \tilde{\gamma}^j, X_t \rangle b_t^j}$, where the b_t^j are distributed $N(0, 1)$, $t = t_1, t_2, \ \ldots \ , t_k, \ \ldots$, and $\tilde{\gamma}^j = (\tilde{\gamma}_1^j, \ \ldots \ , \tilde{\gamma}_N^j)'$.

Equivalently, we suppose that the yield values are observed in additive noise:

$$y_t^j = -\frac{1}{\tau_j} \log F^j(X_t, t) - \frac{\langle \tilde{\gamma}^j, X_t \rangle}{\tau_j} b_t^j \ ,$$

where $1 \leq j \leq m$, $t = t_1, t_2, \ \ldots \ , t_k, \ \ldots \ .$

Now let:

$$\gamma^j = \frac{\tilde{\gamma}^j}{\tau_j} \quad \textit{and}$$

$$\langle g^j, X_t \rangle = -\frac{1}{\tau_j} \log F^j(X_t, t)$$

where:

$$\gamma^j = (\gamma_1^j, \ \ldots \ , \gamma_N^j)' \ ,$$

$$g^j = (g_1^j, \ \ldots \ , g_N^j)' \ , \ \ and$$

$$g_i^j = -\frac{1}{\tau_j} \log\langle e^{-B\tau_j}\underline{1}, e_i\rangle \ ,$$

$1 \leq j \leq m, 1 \leq i \leq N.$

Furthermore, suppose the observation times $t_1 \leq t_2 \leq t_3 \leq \ \ldots$ are equally spaced with $t_{k+1} - t_k = s > 0$. We write $k = t_k$ and let $e^{Cs} = A$. Then we have a discrete time version of the state process with:

$$X_k = AX_{k-1} + M_k$$

where $k \in \mathcal{Z}^+$ and M_k is an $\mathcal{F}_k$-martingale under Q. Also, the observation process $y^j = \{y_k^j\}$ has dynamics:

$$y_k^j = \langle g^j, X_k\rangle + \langle \gamma^j, X_k\rangle v_k^j$$

where $1 \leq j \leq m$ and $k \in \mathcal{Z}^+$. Notice that this is the zero delay observation model of Chapter 2; also, the observation process $y = (y^1, \ \ldots \ , y^m) = \{y_k\}$ is vector-valued. Thus, all the results of Chapter 2, section 2.3.2 carry over to the discussion here.

From the y_k observations, we seek to estimate the state of the chain at a given time, the transition probabilities associated with the chain, and the components of the vectors g^j and γ^j, $1 \leq j \leq m$. Estimators for these parameters are given in Chapter 2, section 2.2.2.

Note that we can also estimate the prices of the varying maturities. For example, with $F^j(X_t, t) = \langle e^{-B\tau_j}\underline{1}, X_t\rangle$, if $k = t_k$ so that $X_{t_k} = X_k$, then:

$$E[F^j(X_k, t_k) \mid \mathcal{Y}_k] = \langle e^{-B\tau_j}\underline{1}, E[X_k \mid \mathcal{Y}_k]\rangle \tag{4.7}$$

where $E[X_k \mid \mathcal{Y}_k] = \frac{q_k}{\langle q_k, \underline{1} \rangle}$ and the ith componet of $e^{-B\tau_j}\underline{1}$ is $\exp\left(-g_i^j \tau_j\right)$, $k \in \mathcal{Z}^+$, $1 \leq j \leq m$, and $1 \leq i \leq N$.

4.3 Application and Results

This section discusses an application of the model to the United States (U.S.) term structure of interest rates. It begins with some general comments.

4.3.1 The Program

A program was written in the **S** language to implement the estimation procedure of Chapter 2, section 2.3.2. A copy of the program is given in Appendix B.

The program uses an iterative procedure. First, it assigns starting values to the model parameters. Then it takes K observations, where each observation is vector-valued with jth component equal to the yield on the jth security, $1 \leq j \leq m$. Using these observations, it estimates parameter values. With these estimates and K further observations, parameter values are re-estimated. The procedure can be repeated any number of times. As mentioned in Chapter 2, the EM algorithm implies that the parameter estimates improve monotonically with each iteration.

4.3.2 The Size of N

As was indicated earlier, N denotes the size of the state space of the Markov chain X. It is the only parameter value that is not estimated in the model; rather, we assign a value to N. In choosing a value for N, we were guided by the same considerations as were enumerated in Chapter 3, section 3.4.1. In the example that follows, N was assigned a value of 4.

4.3.3 Starting Values

Our model:

$$X_k = AX_{k-1} + M_k$$

$$y_k = (y_k^1, \ \ldots \ , y_k^m)$$

$$y_k^j = \langle g^j, X_k \rangle + \langle \gamma^j, X_k \rangle b_k^j$$

where $k \in \mathcal{Z}^+$, $1 \leq j \leq m$, and $g^j, \gamma^j \in \mathbb{R}^N$, has $N^2 + 2mN$ parameters which we seek to estimate and to which we assigned starting values. These parameters are the N^2 elements a_{ij}, $1 \leq i, j \leq N$, of the transition matrix A, the mN elements of the matrix g, and the mN elements of the matrix γ, where the jth columns of g and γ are g^j and γ^j, respectively.

The starting values that were used are discussed in the example that follows.

4.3.4 The Time Step

Because Treasury bill and bond yields may be observed at varying time intervals, for example, weekly or monthly, we sought to allow for this in the analysis. We redefined the ϕ function of Chapter 2, section 2.3.2 to be the $N(0, dt)$ density, where dt represents the time step between observations. The discussion of Chapter 2 had assumed $dt = 1$. In the example that follows, the noise in the observations is, however, assumed to be $N(0, 1)$.

4.3.5 Checking the Model

The equation:

$$\begin{aligned} y_k^j &= \langle g^j, X_k \rangle + \langle \gamma^j, X_k \rangle b_k^j \\ &= g^j(X_k) + \gamma^j(X_k) b_k^j \end{aligned}$$

where $\{b_k^j\}$ is a sequence of i.i.d. $N(0,1)$ random variables, $k \in \mathcal{Z}^+$, and $1 \leq j \leq m$, suggests that the statistic:

$$\frac{y_k - g^j(\hat{X}_k)}{\gamma^j(\hat{X}_k)} ,$$

where $\hat{X}_k = \frac{q_k}{\langle q_k, 1 \rangle}$, should be normally distributed with mean zero and variance one. In the example that follows, we check the model by testing for the normality of this statistic and by calculating its mean and variance.

4.3.6 Scaling the Yield Values

We observed in Chapter 2, section 2.3.2 that:

$$\begin{aligned} \Gamma^i(y_k) &= \Gamma^i(y_k^1, \ \ldots \ , y_k^m) \\ &= \prod_{j=1}^{m} \frac{\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right)}{\gamma_i^j \phi(y_k^j)} \end{aligned}$$

where $k \in \mathcal{Z}^+$, $1 \leq i \leq N$, and ϕ denotes the $N(0,1)$ density. For small γ_i^j and large $|y_k^j - g_i^j|$, $\Gamma^i(y_k)$ goes to zero and parameter estimates are not obtained. To avoid this, we specified a tolerance value, usually 1.0e-10, below which we would not let the value of $\Gamma^i(y_k)$ fall. However, if the values of y_k^j, g_i^j, and γ_i^j were such that $\Gamma^i(y_k)$ assumed this tolerance value for most values of i, j, and k, important yield information was lost to the subsequent parameter estimation.

In the application that follows, we explored the usefulness of multiplying the y_k^j values by a factor $\alpha > 1$. By scaling the yields in this way, we hoped to obtain larger $\Gamma^i(y_k)$ values.

It should be noted that if we multiply the y_k^j by $\alpha > 1$, we must, of course, adjust the parameter estimates. If we consider the model:

$$y_k^j = \langle g^j, X_k \rangle + \langle \gamma^j, X_k \rangle b_k^j ,$$

then with:

$$\begin{aligned} Y_k^j &= \alpha y_k^j \\ &= \alpha\langle g^j, X_k\rangle + \alpha\langle \gamma^j, X_k\rangle b_k^j \\ &= \langle \alpha g^j, X_k\rangle + \langle \alpha\gamma^j, X_k\rangle b_k^j \\ &= \langle g^j(\alpha), X_k\rangle + \langle \gamma^j(\alpha), X_k\rangle b_k^j \ , \end{aligned}$$

we have $g^j(\alpha) = \alpha g^j$ and $\gamma^j(\alpha) = \alpha\gamma^j$. Therefore, $\widehat{g^j(\alpha)} = \alpha\widehat{g^j}$ and $\widehat{g^j} = \frac{\widehat{g^j(\alpha)}}{\alpha}$. Similarly, $\widehat{\gamma^j} = \frac{\widehat{\gamma^j(\alpha)}}{\alpha}$.

4.3.7 Application of the Model

In the application of the model considered here, we analyzed a data set consisting of 225 weekly observations on the yields of 1-month, 3-month, and 6-month U.S. Treasury bills and 2-year, 5-year, 7-year, 10-year, and 30-year U.S. bonds. The sample period ran from January 17, 1992 to June 1, 1996. The data were compiled by the Royal Bank of Canada and published in *The Financial Post.*

In the analysis, we processed the U.S. yield data in 15 groups of 15 yield vectors each. At the end of each pass through the data, parameter estimates were updated using the formulas of Chapter 2, section 2.2.2. We let N, the size of the state space of the Markov chain X, equal four; dt, the time step between observations, be 5/252; and the vector, $\tau = (\tau_1, \ \ldots \ , \tau_m)'$, of terms to maturity of the various securities be $(21/252, 63/252, 126/252, 2, 5, 7, 10, 30)'$. Here, τ_j denotes the term to maturity of the jth security, $1 \leq j \leq m$, and $m = 8$.

Scaling the Yield Values, Starting Values, Parameter Re-estimation, and Checking the Model:

In the analysis, we let the scaling factor, α, equal 1.0, 2.5, 5.0, 7.5, 10.0, and 20.0, in turn. In each case, we multiplied the y_k^j values, $1 \leq j \leq 8$, $1 \leq k \leq 15$, by α in each pass through the data. We also adjusted the starting value of the g matrix according to the value of α in order to make $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$, large in the first pass through the data. We then observed:

1. for all values of α used, $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$, assumed the tolerance value of 1.0e-10 in the second and all subsequent passes, and

2. the re-estimated values of the elements of the γ matrix were larger for larger values of α.

In choosing the value of α that we judged to give the best results, we considered the γ_i^j values, $1 \leq i \leq 4$, $1 \leq j \leq 8$, since larger γ_i^j values contributed to larger values of $\Gamma^i(y_k)$ in the subsequent pass through the data. We also considered the mean and variance of the test statistic:

$$\frac{y_k^j - g^j(\hat{X}_k)}{\gamma^j(\hat{X}_k)} \, , 1 \leq j \leq 8, 1 \leq k \leq 15 \, .$$

Based on these considerations, we concluded that the results for $\alpha = 5.0$ were superior to those obtained in the other five cases.

For the case $\alpha = 5.0$, the hypothesis of normality, in regard to the distribution of the test statistic, was not rejected 109 times, out of a possible total of 120, at a significance level of 0.05. The hypothesis of zero mean was rejected for all securities. The hypothesis of unit variance met with mixed results; for example, for the 1-, 3-, and 6-month Treasury bills and the 2-, 5-, 7-, 10-, and 30-year bonds,

the 95 percent confidence intervals for the sample variance were, respectively, (2.5112, 7.2324), (1.3448, 6.2199), (1.0624, 7.1345), (1.3096, 4.8475), (0.7910, 1.8919), (0.6054, 1.3604), (0.4224, 0.9877), and (0.1901, 0.5600).

For the case $\alpha = 5.0$, Table 4.1 reports the results for the first four securities. It gives the starting values that were assigned to the distribution of the state of the Markov chain that is, to $E[X_0]$, the transition matrix A, and the g matrix. Note that all the entries in the γ matrix were assigned a starting value of 0.25. Table 4.1 also gives the re-estimated values of the parameters after the fifteenth pass through the data. Estimated Treasury bill and bond prices are also reported. Notice that the columns of the g and γ matrices and the price vector correspond, respectively, to the 1-, 3-, and 6-month Treasury bills and the 2-year bond. Table 4.2 is similarly structured and gives results for the 5-, 7-, 10-, and 30-year bonds.

Yield Estimation:

At the end of each pass through the data and for each security, we calculated the yield vector corresponding to the estimated price vector using the formula:

$$P = e^{-rt}$$

or

$$r = \frac{\log P}{-t}$$

where P denotes the treasury bill or bond price, r the corresponding yield, and t denotes time to maturity. Figures 4.1 to 4.8 provide plots of actual yields and estimated yields for each of the securities in turn.

We also regressed actual yields on estimated yields, using the model:

$$Actual\ yield = \alpha + \beta * Estimated\ yield + \epsilon\ .$$

The regression results obtained were assessed on the basis of the following criteria proposed by Fama and Gibbons [11]: (1) conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one; (2) serially uncorrelated residuals; and (3) a low residual standard error. Table 4.3 reports the results for the 1-, 3-, and 6-month Treasury bills and the 2-year bond. Table 4.4 gives results for the 5-, 7-, 10-, and 30-year bonds. For each of the securities, we can conclude that the intercept is zero and the slope is one. With the exception of the 3-month Treasury bill, we can also conclude that the residuals do not display first-order serial correlation.

4.3.8 Concluding Remarks

We have observed that, for the case $\alpha = 5.0$ and for all other values of α considered, $\Gamma^i(y_k)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$, assumed the tolerance value of $1.0e\text{-}10$ in the second and all subsequent passes through the data. This is undesirable because the information in the data enters the parameter estimation in large part through the $\Gamma^i(y_k)$ values. As a result, we gave some thought to how the $\Gamma^i(y_k)$ values might be increased. Based in part on our analysis of Chapter 3, we thought that by adjusting the units for measuring time we might be able to increase the $\Gamma^i(y_k)$ values. Recall that:

$$\begin{aligned} \Gamma^i(y_k) &= \Gamma^i(y_k^1, \ \ldots \ , y_k^m) \\ &= \prod_{j=1}^{m} \frac{\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right)}{\gamma_i^j \phi(y_k^j)} , \end{aligned} \tag{4.8}$$

$k \in \mathcal{Z}^+$, $1 \leq i \leq N$, where:

$$\phi(x) = (2\pi dt)^{-\frac{1}{2}} \exp\left(\frac{-x^2}{2dt}\right)$$

is the $N(0, dt)$ density. In the example we are considering here, $dt = 5/252$. For the values of x observed, $\exp\left(\frac{-x^2}{2dt}\right)$ tended to be small, near zero, for $dt = 5/252$. Therefore, we thought that if we measured time in days for example, with $dt = 5$, we might get larger $\Gamma^i(y_k)$ values. However, formula (4.7) for the expected price of the jth security, $1 \leq j \leq m$, requires that time be measured in years. Thus this avenue for increasing the $\Gamma^i(y_k)$ values was not available.

Also, we observe in equation (4.8) that $\Gamma^i(y_k)$ is a product. It follows that if, for a given i, $1 \leq i \leq 4$, and a given k, $1 \leq k \leq 15$, the numbers, $\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right) / \gamma_i^j \phi(y_k^j)$, $1 \leq j \leq 8$, are small (less than one), then $\Gamma^i(y_k)$ is even smaller. We decided to determine the values of these numbers by examining each of the securities individually, that is, by employing our model with $m = 1$. When we did this, we observed that for each j, $1 \leq j \leq 8$, the numbers $\phi\left(\frac{y_k^j - g_i^j}{\gamma_i^j}\right) / \gamma_i^j \phi(y_k^j)$, $1 \leq i \leq 4$, $1 \leq k \leq 15$, assumed the tolerance value of 1.0e-10 in the second and all subsequent passes through the data, for all values of α considered when $dt = 5/252$. Thus the problem of low $\Gamma^i(y_k)$ values is one that is not easily overcome in this model.

TABLE 4.1: **Starting values and parameter estimates – 1-, 3-, and 6-month Treasury bills and 2-year bond**

Starting values:

$E[X_0]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:
$$\begin{pmatrix} 0.23 & 0.27 & 0.20 & 0.30 \\ 0.20 & 0.30 & 0.30 & 0.27 \\ 0.30 & 0.20 & 0.27 & 0.20 \\ 0.27 & 0.23 & 0.23 & 0.23 \end{pmatrix}$$

g matrix:
$$\begin{pmatrix} 0.30 & 0.30 & 0.30 & 0.30 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.29 & 0.29 & 0.29 & 0.29 \\ 0.27 & 0.27 & 0.27 & 0.27 \end{pmatrix}$$

After the fifteenth pass:

$E[X_{15} \mid \mathcal{Y}_{15}]$ vector: (0.32311655, 0.01147894, 0.46589404, 0.19951047)

A matrix:
$$\begin{pmatrix} 0.33987850 & 0.3089177 & 0.2945645 & 0.3634610 \\ 0.00591622 & 0.0173240 & 0.0131932 & 0.0161484 \\ 0.48277644 & 0.3982955 & 0.4951553 & 0.3741108 \\ 0.17142882 & 0.2754626 & 0.1970868 & 0.2462796 \end{pmatrix}$$

g matrix:
$$\begin{pmatrix} 0.0499820 & 0.0495271 & 0.0502158 & 0.0585971 \\ 0.0491129 & 0.0486889 & 0.0487352 & 0.0543676 \\ 0.0500093 & 0.0495562 & 0.0502704 & 0.0587449 \\ 0.0499311 & 0.0494756 & 0.0501196 & 0.0583332 \end{pmatrix}$$

γ matrix:
$$\begin{pmatrix} 0.0080009 & 0.0078514 & 0.0080750 & 0.0110269 \\ 0.0077240 & 0.0075892 & 0.0076104 & 0.0095375 \\ 0.0080096 & 0.0078604 & 0.0080920 & 0.0110781 \\ 0.0079848 & 0.0078353 & 0.0080451 & 0.0109356 \end{pmatrix}$$

Estimated prices: (0.9958441, 0.9876961, 0.9752099, 0.8894703)

TABLE 4.2: **Starting values and parameter estimates – 5-, 7-, 10-, and 30-year bonds**

Starting values:

$E[X_0]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:
$$\begin{pmatrix} 0.23 & 0.27 & 0.20 & 0.30 \\ 0.20 & 0.30 & 0.30 & 0.27 \\ 0.30 & 0.20 & 0.27 & 0.20 \\ 0.27 & 0.23 & 0.23 & 0.23 \end{pmatrix}$$

g matrix:
$$\begin{pmatrix} 0.30 & 0.30 & 0.30 & 0.30 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.29 & 0.29 & 0.29 & 0.29 \\ 0.27 & 0.27 & 0.27 & 0.27 \end{pmatrix}$$

After the fifteenth pass:

$E[X_{15} \mid \mathcal{Y}_{15}]$ vector: (0.32311655, 0.01147894, 0.46589404, 0.19951047)

A matrix:
$$\begin{pmatrix} 0.33987850 & 0.3089177 & 0.2945645 & 0.3634610 \\ 0.00591622 & 0.0173240 & 0.0131932 & 0.0161484 \\ 0.48277644 & 0.3982955 & 0.4951553 & 0.3741108 \\ 0.17142882 & 0.2754626 & 0.1970868 & 0.2462796 \end{pmatrix}$$

g matrix:
$$\begin{pmatrix} 0.0621530 & 0.0640264 & 0.0648942 & 0.0676422 \\ 0.0580654 & 0.0604820 & 0.0618388 & 0.0656157 \\ 0.0622985 & 0.0641537 & 0.0650050 & 0.0677177 \\ 0.0618951 & 0.0638003 & 0.0646984 & 0.0675094 \end{pmatrix}$$

γ matrix:
$$\begin{pmatrix} 0.0124077 & 0.0131541 & 0.0135099 & 0.0146585 \\ 0.0108665 & 0.0117646 & 0.0122835 & 0.0137987 \\ 0.0124616 & 0.0132033 & 0.0135538 & 0.0146902 \\ 0.0123120 & 0.0130667 & 0.0134322 & 0.0146026 \end{pmatrix}$$

Estimated prices: (0.7329999, 0.6389073, 0.5227190, 0.09252945)

TABLE 4.3: **Regressions of actual yields on estimated yields – 1-, 3-, and 6-month Treasury bills and 2-year bond**

Parameter	Term to maturity of security			
	1-month	3-month	6-month	2-year
α	0.000307	0.000595	0.000148	-0.002347
	(0.002543)	(0.002528)	(0.003160)	(0.005246)
β	0.99864	0.99688	1.00277	1.05136
	(0.06271)	(0.05967)	(0.07141)	(0.09795)
R-squared	0.951	0.955	0.938	0.899
Durbin-Watson D statistic	2.08	1.34	1.53	2.00
s	0.002445	0.002364	0.002914	0.003875

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 4.4: **Regressions of actual yields on estimated yields 5-, 7-, 10-, and 30-year bonds**

Parameter	Term to maturity of security			
	5-year	7-year	10-year	30-year
α	-0.003115	-0.000768	0.000112	0.003958
	(0.006406)	(0.006336)	(0.006973)	(0.006305)
β	1.0517	1.00764	0.9960	0.93919
	(0.1048)	(0.09989)	(0.1055)	(0.08903)
R-squared	0.886	0.887	0.873	0.895
Durbin-Watson D statistic	1.94	1.97	1.47	1.56
s	0.003162	0.002797	0.002738	0.002044

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

FIGURE 4.1 Yields on 1-month T-bill

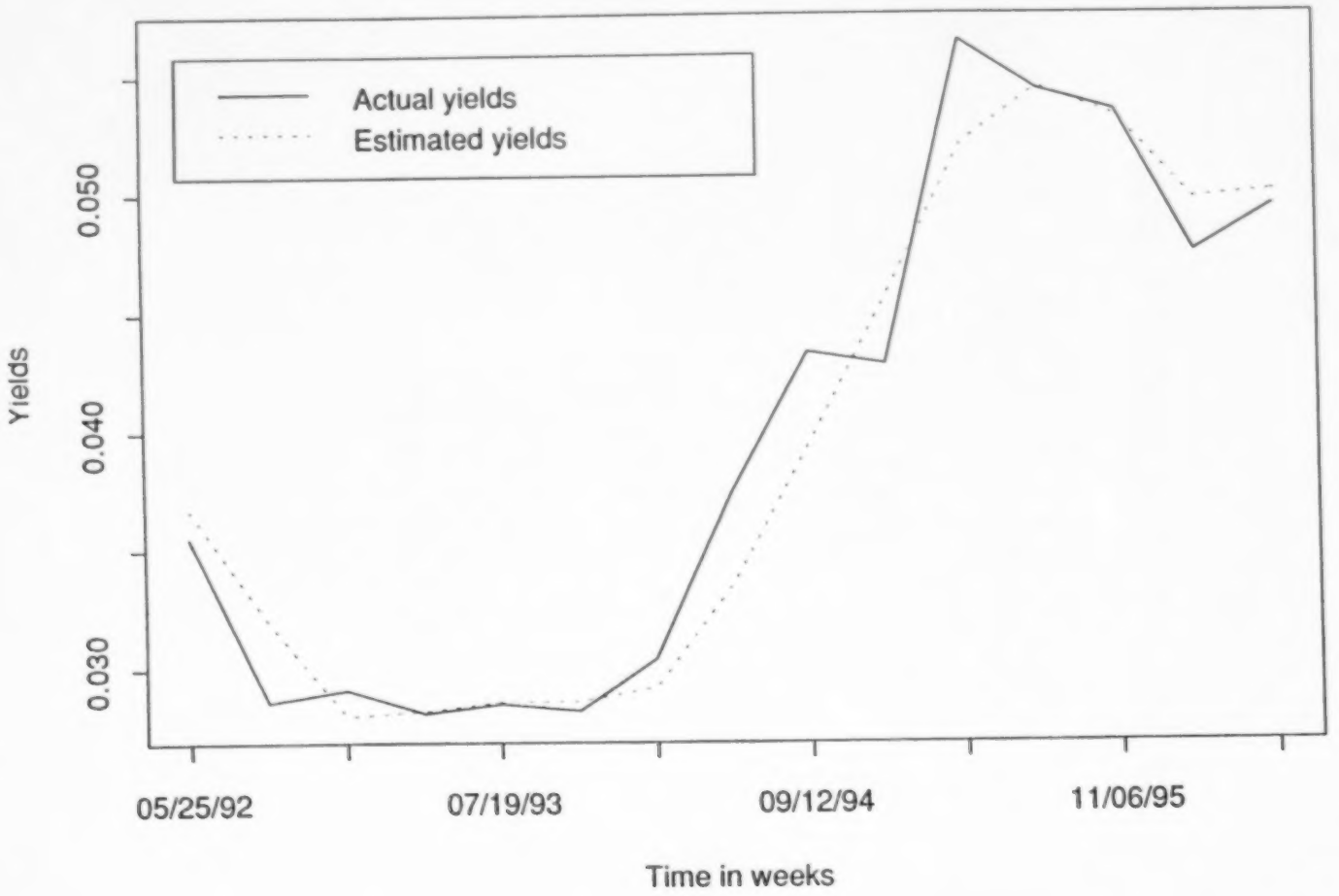


FIGURE 4.2 Yields on 3-month T-bill

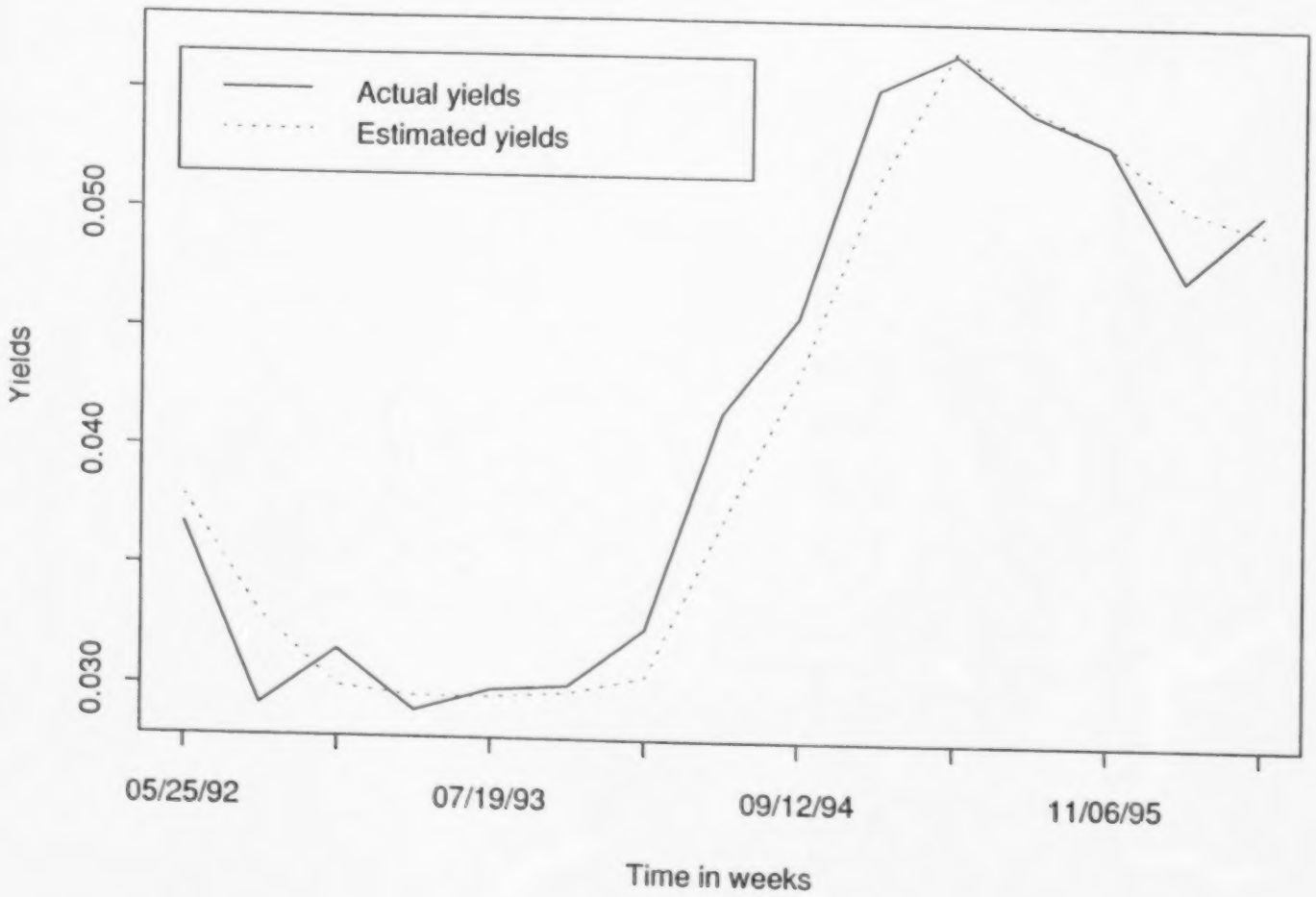


FIGURE 4.3 Yields on 6-month T-bill

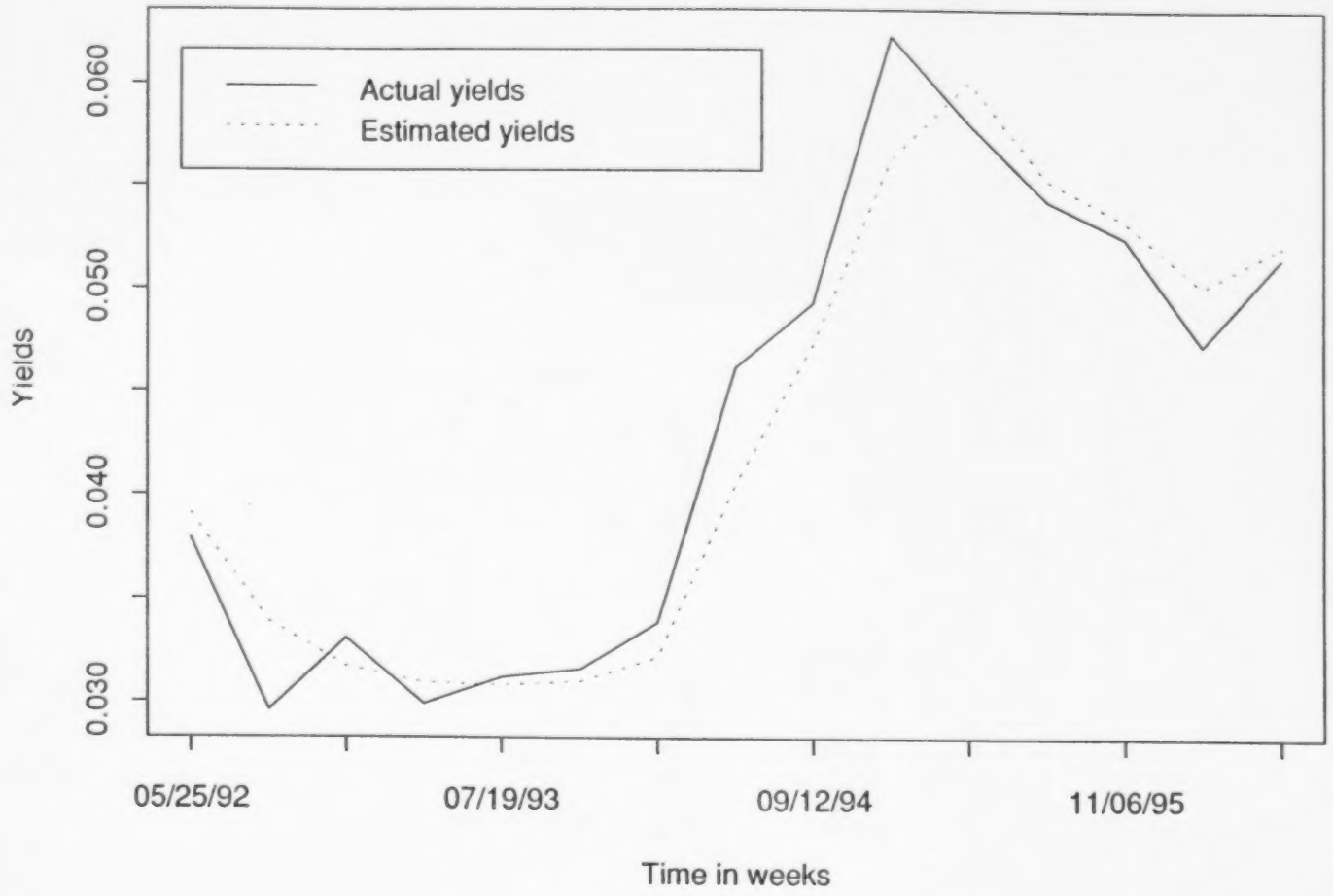


FIGURE 4.4 Yields on 2-year bond

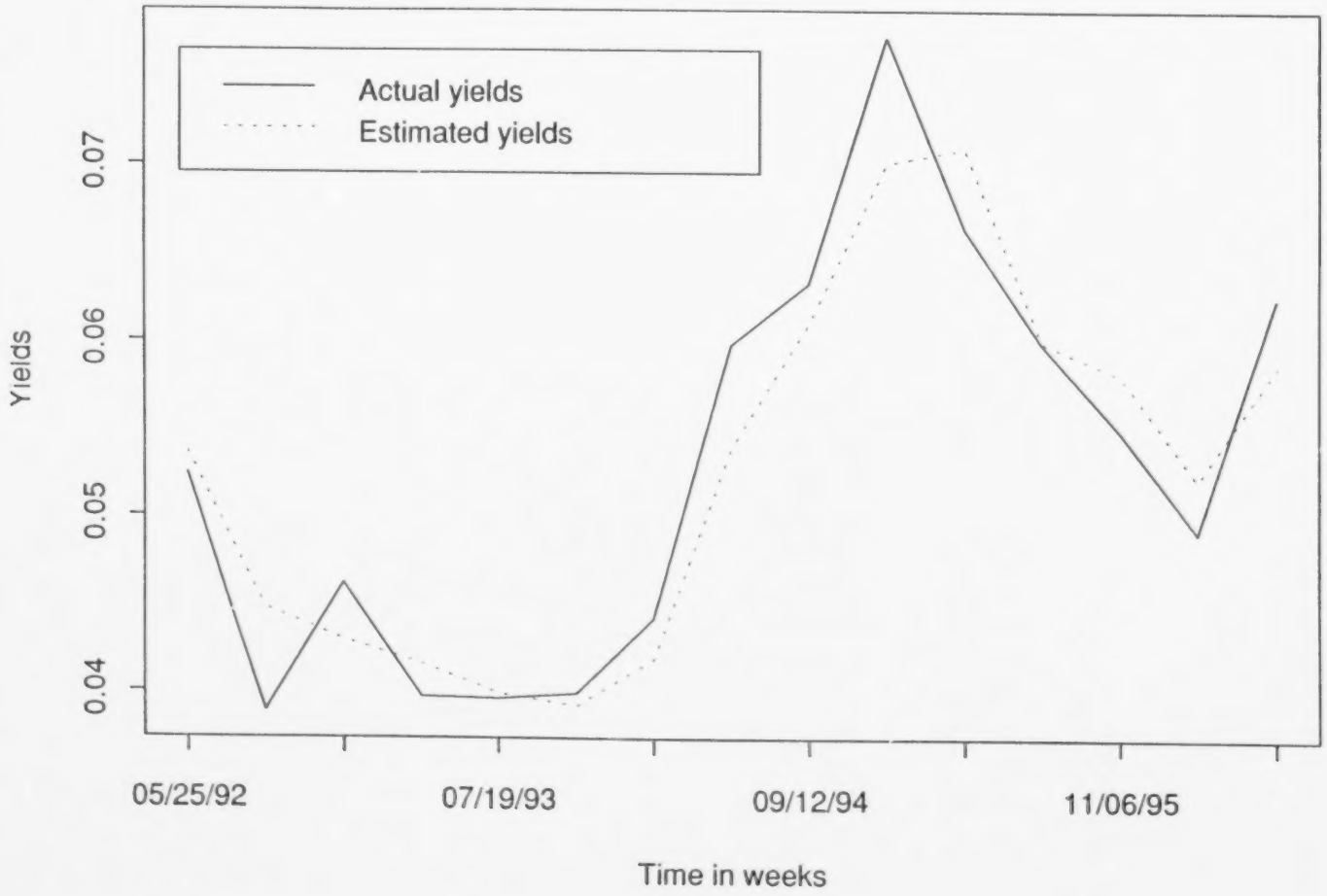


FIGURE 4.5 Yields on 5-year bond

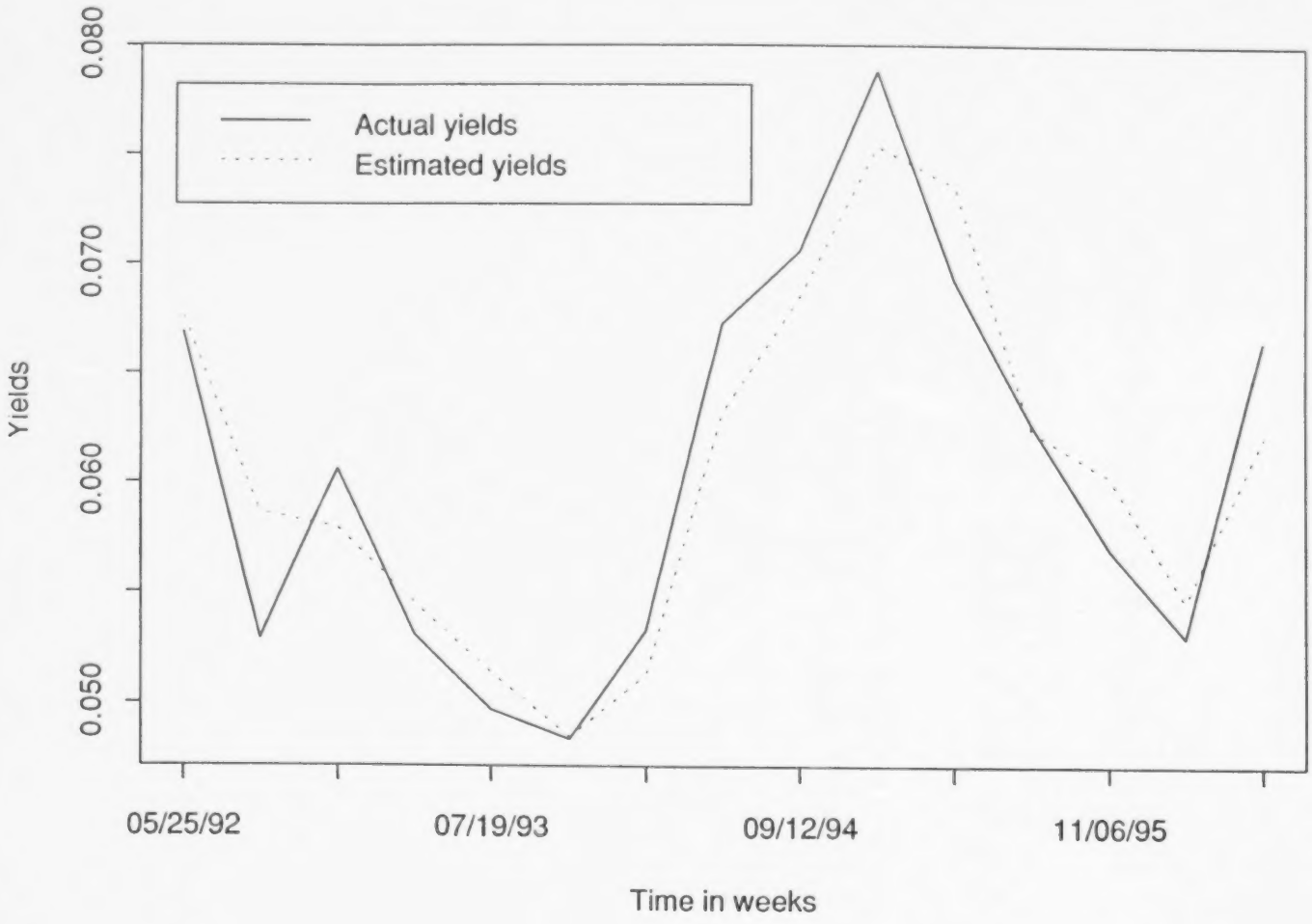


FIGURE 4.6 Yields on 7-year bond

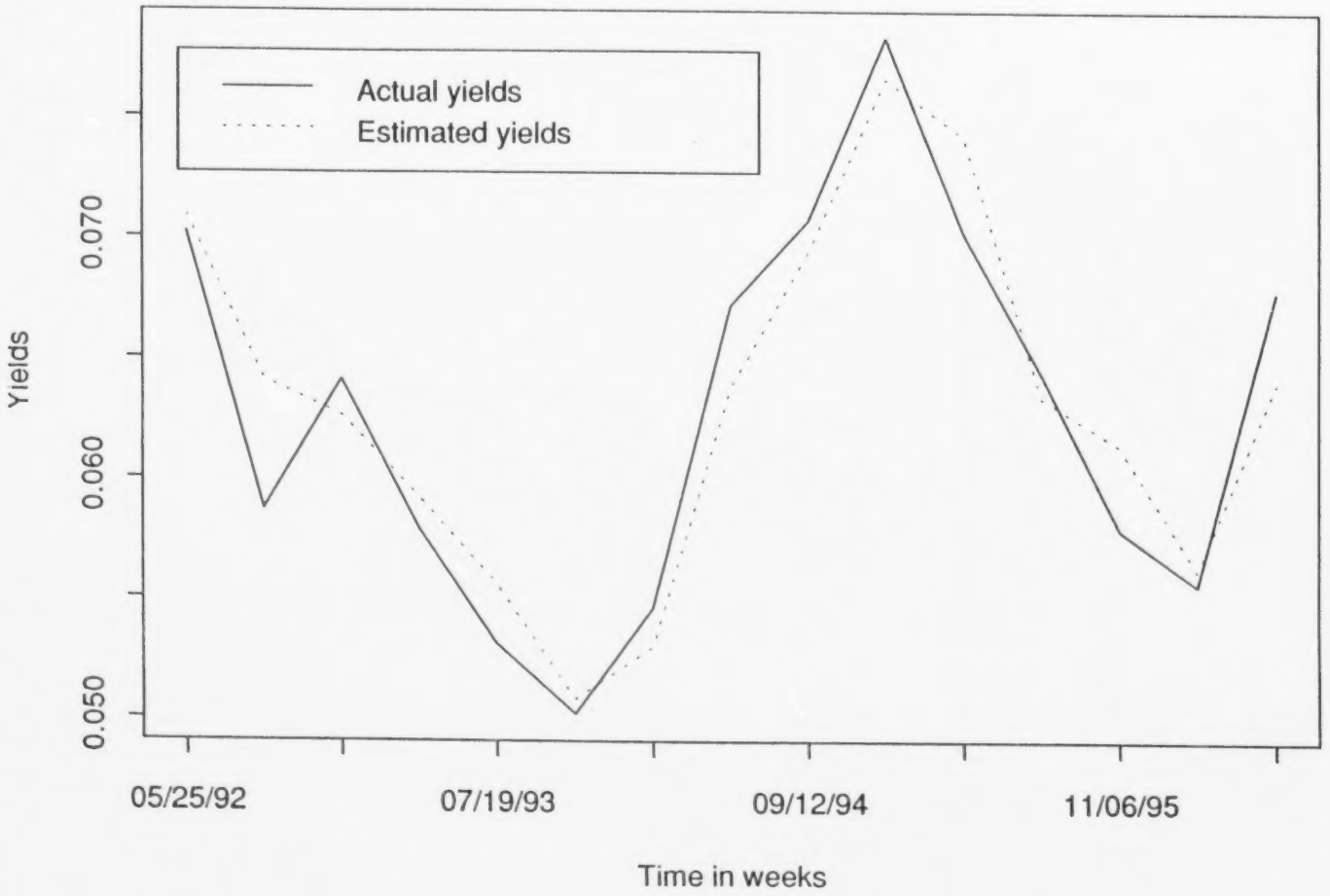


FIGURE 4.7 Yields on 10-year bond

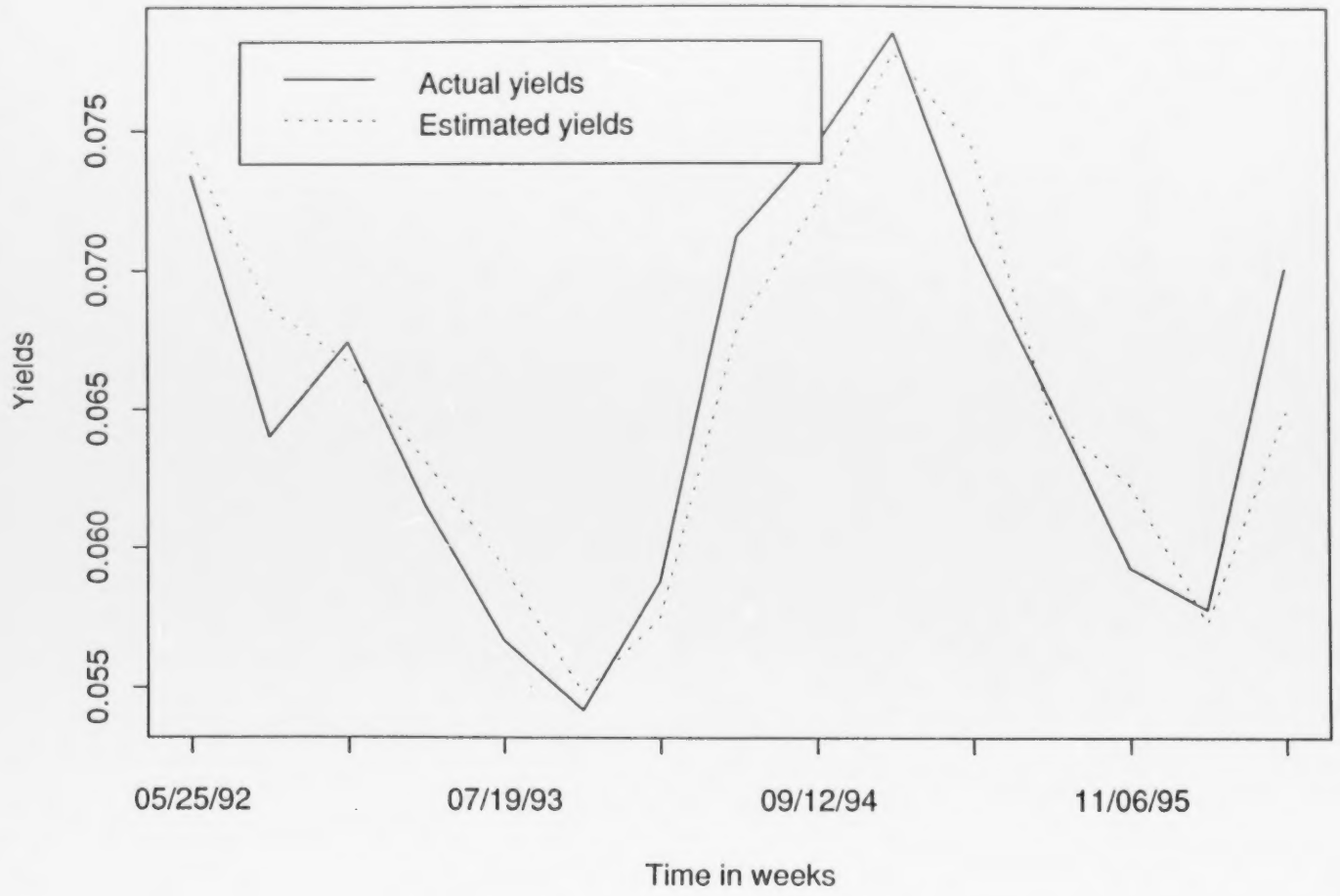
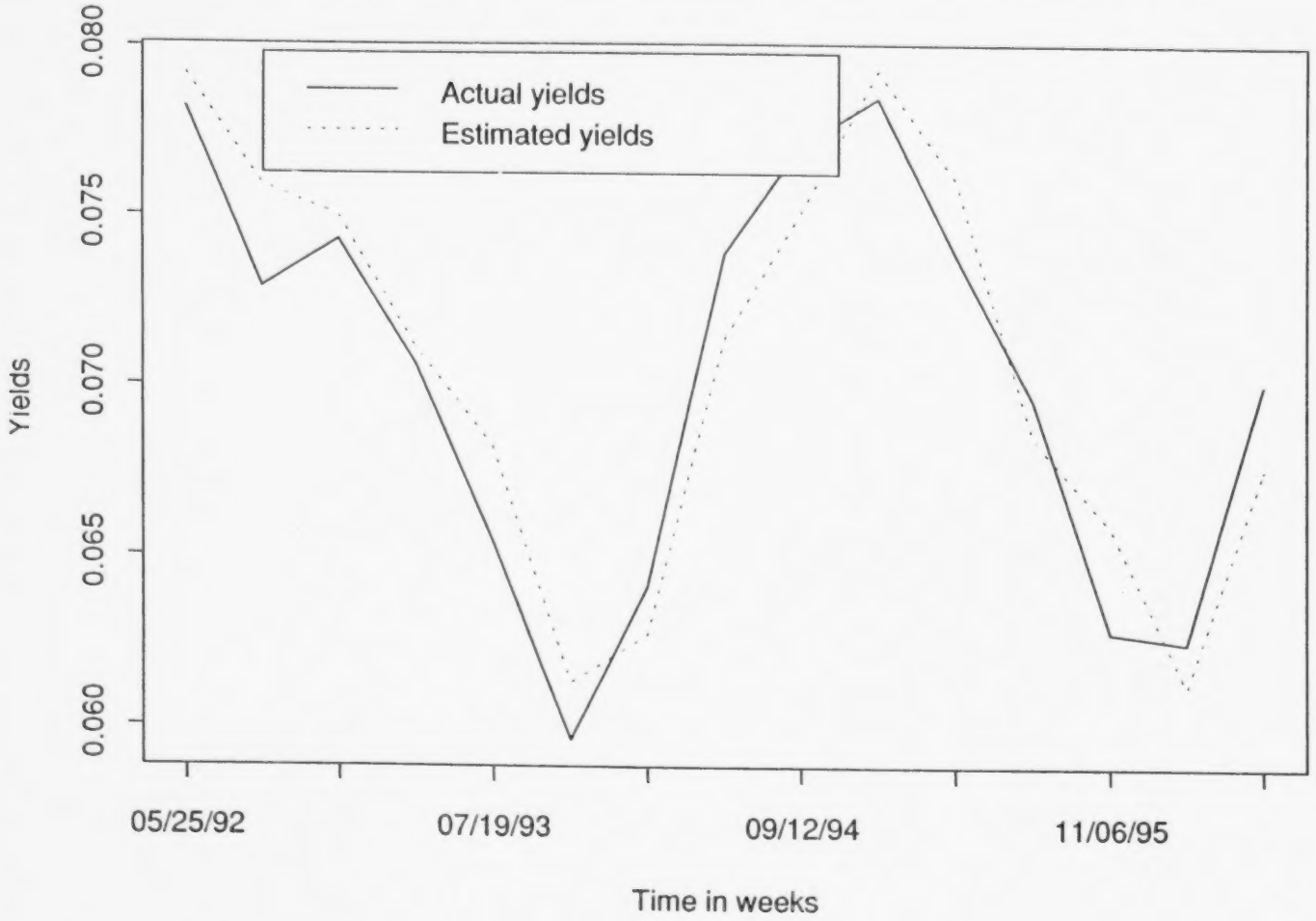


FIGURE 4.8 Yields on 30-year bond



Bibliography

[1] Philippe Artzner and Freddy Delbaen. Term structure of interest rates: The martingale approach. *Advances in Applied Mathematics*, 10:95–129, 1989.

[2] Patrick Billingsley. *Probability and Measure*. John Wiley and Sons, New York, 1986.

[3] John C. Cox, Jonathan E. Ingersoll Jr., and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

[4] Darrell Duffie and Rui Kan. Multi-factor term structure models. A survey designed for presentation at a meeting of the Royal Society in London, November 10-11, 1993.

[5] R.J. Elliott. *Stochastic Calculus and Applications*. Springer-Verlag, New York, 1982.

[6] R.J. Elliott. New finite-dimensional filters and smoothers for noisily observed Markov chains. *IEEE Transactions on Information Theory*, 39:265–271, 1993.

[7] R.J. Elliott. Exact adaptive filters for Markov chains observed in Gaussian noise. *Automatica*, 30:1399–1408, 1994.

[8] R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models: Estimation and Control*, volume 29 of *Applications of Mathematics*. Springer-Verlag, New York, 1995.

[9] R.J. Elliott and W.C. Hunter. Filtering a discrete time process. The addresses of the authors are, respectively, the Department of Mathematical Sciences, University of Alberta and the Federal Reserve Bank, Chicago.

[10] R.J. Elliott, W.C. Hunter, and B.M. Jamieson. Drift and volatility estimation in discrete time. The address of R.J. Elliott and B.M. Jamieson is the Department of Mathematical Sciences, University of Alberta. The address of W.C. Hunter is the Federal Reserve Bank, Chicago.

[11] R.J. Elliott, W.C. Hunter, and B.M. Jamieson. Financial signal processing. The address of R.J. Elliott and B.M. Jamieson is the Department of Mathematical Sciences, University of Alberta. The address of W.C. Hunter is the Federal Reserve Bank, Chicago.

[12] Eugene F. Fama and Michael R. Gibbons. A comparison of inflation forecasts. *Journal of Monetary Economics*, 13:327–348, 1984.

[13] J.M. Harrison and D.M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

[14] J.M. Harrison and S.R. Pliska. Martingale and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260, 1981.

[15] Vikram Krishnamurthy and John B. Moore. On-line estimation of hidden Markov model paramenters based on the Kullback-Leibler information measure. *IEEE Transactions on Signal Processing*, 41:2557–2573, 1993.

[16] Damien Lamberton and Bernard Lapeyre. *Introduction au Calcul Stochastique Appliqué à la Finance.* ellipses, Paris, 1991.

[17] Inc. Minitab. *Minitab Reference Manual Release 7.* Minitab, Inc., 1989.

[18] Bernt Oksendal. *Stochastic Differential Equations.* Springer-Veralg, Berlin, fourth edition, 1995.

[19] Neil D. Pearson and Tong-Sheng Sun. Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model. *The Journal of Finance*, 49:1279–1304, 1994.

[20] Paul Samuelson. Mathematics of speculative price. *Industrial Management Review*, 6:13–39, 1965.

Appendix A

Computer Program for Chapter 3

This appendix contains a copy of the **S** program written to implement the estimation procedure of Chapter 3.

```
# The parameter values that must be entered are: No.batches, dt, m,
# n, TOLERANCE, and factor.  Also enter the dimensions of the data
# matrix "prices".

# Enter the data and parameter values.

# The data are in one column in a file called "price".
# Read the data into one column of a data frame called "prc".
# Then read the data into a matrix called "prices". Specify the
# dimensions of "prices".
# The number of columns of "prices" equals the number of batches
# (equivalent to the number of passes).
# The number of rows of "prices" equals the number of prices in each
# batch.

prc<-read.table("price")
prices<-matrix(prc[ ,1],16,2)

No.batches<-2
cat("\n","The number of price batches is:","\n","\n")
print(No.batches)

# Enter the time step.

dt<-21  # Number of working days between price quotes
cat("\n","The time step is:","\n","\n")
print(dt)

m<-2
n<-2
N<-m*n
cat("\n","The value of m is:","\n","\n")
print(m)
cat("\n","The value of n is:","\n","\n")
print(n)
cat("\n","The value of N is:","\n","\n")
print(N)

TOLERANCE<-(1.0e-10) # Elements in the GAMMA matrix are not allowed
# to fall below this level.
cat("\n","The TOLERANCE value is:","\n","\n")
print(TOLERANCE)

factor<-2.5
cat("\n","The factor multiplying the y values is:","\n","\n")
print(factor)

# Initialize.

g<-c(0.9,0.1,0.8,0.2)
cat("\n","The g vector is:","\n","\n")
print(g)

gamma<-rep(0.25,N)
cat("\n","The gamma vector is:","\n","\n")
print(gamma)
```

```
A<-matrix((1/N),nrow=N,ncol=N)
cat("\n","The A matrix is:","\n","\n")
print(A)

# Run the batches of data.

for (b in 1:No.batches) {

cat("\n","The batch number is:","\n","\n")
print(b)

cat("\n","The prices in this batch are:","\n","\n")
print(prices[ ,b])

y<-vector(length=(length(prices[ ,b])-1))
        for (i in 1:(length(y)))
        y[i]<-(factor*log(prices[ ,b][i+1]/prices[ ,b][i]))
cat("\n","The y vector is:","\n","\n")
print(y)
K<-length(y)
cat("\n","The mean of the y vector is:","\n","\n")
print(mean(y))
cat("\n","The variance of the y vector is:","\n","\n")
print(var(y))

phi<-function(input) {
        (((2*pi*dt)^((-1)/2))*exp(((-1)*((input)^(2)))/(2*dt))) }

GAMMA<-matrix(nrow=K,ncol=N)
        for (i in 1:K) {
        for (j in 1:N) {
        GAMMA[i,j]<-(phi((y[i]-g[j])/gamma[j])/
        (gamma[j]*phi(y[i])))
        if (GAMMA[i,j]<TOLERANCE) GAMMA[i,j]<-TOLERANCE }}
cat("\n","The GAMMA matrix is:","\n","\n")
print(GAMMA)

GAMMA.A<-array(dim=c(N,K,N))
        for (i in 1:K){
        for (j in 1:N)
        GAMMA.A[ ,i,j]<-GAMMA[i,j]*A[ ,j] }

# Recurrences.

e.matrix<-diag(N)

# Get q vectors.

q0<-rep((1/N),N)

q.matrix<-matrix(0,nrow=N,ncol=(K+1)) # This is used to store the
# q vector after each iteration.
q.matrix[ ,1]<-q0

q.norm<-matrix(0,nrow=N,ncol=K) # This is used to store the
# normalized q vector after each iteration. This is our estimate
```

```
# of E[Xk/Yk], for k= 1, 2, ..., K.

for (k in 1:K) {

        q1<-matrix(nrow=N,ncol=N)
                for (i in 1:N)
                q1[ ,i]<-(q0[i]*GAMMA.A[ ,k,i])

        rm(q0)

        q<-vector(length=N)
                for (i in 1:(length(q)))
                q[i]<-sum(q1[i, ])

        rm(q1)

        q.matrix[ ,(k+1)]<-q  # We store q at the end of each
        # iteration.

        q0<-q

        q.norm[ ,k]<-(q.matrix[ ,(k+1)]/sum(q.matrix[ ,(k+1)])) )

        cat("\n","The q vector after K iterations is:","\n","\n")
        print(q0)
        cat("\n","The normalized q vector, our estimate of E[XK/YK],")
        cat("\n","is:","\n","\n")
        print(q.norm[ ,K])

        cat("\n","The denominator of q.norm[ ,K], i.e. of our")
        cat("\n","estimate of E[XK/YK],is:")
        cat("\n","\n")
        q.sum<-(sum(q.matrix[ ,K+1]))
        print(q.sum)

# Get N arrays.

N0<-array(0,dim=c(N,N,N))

for (k in 1:K) {

        N1<-array(dim=c(N,N,N,N))
                for (r in 1:N) {
                for (s in 1:N) {
                for (i in 1:N)
                N1[ ,r,s,i]<-N0[ ,r,s][i]*GAMMA.A[ ,k,i] ))

        rm(N0)

        N.hat<-array(dim=c(N,N,N))
                for (i in 1:N) {
                for (r in 1:N) {
                for (s in 1:N) {
                N.hat[i,r,s]<-if (i==s) (sum(N1[i,r,s, ])+
                (q.matrix[ ,k][r]*A[s,r]*GAMMA[k,r]))
                else (sum(N1[i,r,s, ])) )))
```

```
rm(N1)

        N0<-N.hat }
        rm(A,GAMMA,e.matrix)

        cat("\n","The N.hat array after K iterations is:","\n","\n")
        print(N0)

# Get J matrices.

J0<-matrix(0,nrow=N,ncol=N)

for (k in 1:K) {

        J1<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (i in 1:N)
                J1[ ,r,i]<-J0[ ,r][i]*GAMMA.A[ ,k,i] }

        rm(J0)

        J<-matrix(nrow=N,ncol=N)
                for (i in 1:N) {
                for (r in 1:N)
                J[i,r]<-(sum(J1[i,r, ])+
                (q.matrix[ ,k][r]*GAMMA.A[i,k,r])) }

        rm(J1)

        J0<-J }

        cat("\n","The J matrix after K iterations is:","\n","\n")
        print(J0)

# Get Gy matrices.

G0<-matrix(0,nrow=N,ncol=N)

for (k in 1:K) {

        Gy1<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (i in 1:N)
                Gy1[ ,r,i]<-G0[ ,r][i]*GAMMA.A[ ,k,i] }

        rm(G0)

        Gy<-matrix(nrow=N,ncol=N)
                for (i in 1:N) {
                for (r in 1:N)
                Gy[i,r]<-(sum(Gy1[i,r, ])+
                (q.matrix[ ,k][r]*y[k]*GAMMA.A[i,k,r])) }

        rm(Gy1)

        G0<-Gy }
```

```
cat("\n","The Gy matrix after K iterations is:","\n","\n")
        print(G0)

# Get the Gysq matrices.

G0.sq<-matrix(0,nrow=N,ncol=N)

for (k in 1:K) {

        Gysq1<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (i in 1:N)
                Gysq1[ ,r,i]<-G0.sq[ ,r][i]*GAMMA.A[ ,k,i] }

        rm(G0.sq)

        Gysq<-matrix(nrow=N,ncol=N)
                for (i in 1:N) {
                for (r in 1:N)
                Gysq[i,r]<-(sum(Gysq1[i,r, ])+
                (q.matrix[ ,k][r]*(y[k]^2)*GAMMA.A[i,k,r])) }

        rm(Gysq1)

        G0.sq<-Gysq }

        cat("\n","The Gysq matrix after K iterations is:","\n","\n")
        print(G0.sq)

# Get the totals.

sigma.N<-matrix(nrow=N,ncol=N)
        for (r in 1:N) {
        for (s in 1:N)
        sigma.N[r,s]<-sum(N.hat[ ,r,s]) }

cat("\n","Sigma.N is:","\n","\n")
print(sigma.N)

N.check<-matrix(nrow=N,ncol=N)
        for (r in 1:N){
        for (s in 1:N)
        N.check[r,s]<-(sigma.N[r,s]/q.sum))

cat("\n","The N.check matrix is:","\n","\n")
print(N.check)

sigma.J<-vector(length=N)
        for (r in 1:(length(sigma.J)))
        sigma.J[r]<-sum(J[ ,r])

cat("\n","Sigma.J is:","\n","\n")
print(sigma.J)

sigma.Gy<-vector(length=N)
        for (r in 1:(length(sigma.Gy)))
```

```
sigma.Gy[r]<-sum(Gy[ ,r])

cat("\n","Sigma.Gy is:","\n","\n")
print(sigma.Gy)

sigma.Gysq<-vector(length=N)
        for (r in 1:(length(sigma.Gysq)))
        sigma.Gysq[r]<-sum(Gysq[ ,r])

cat("\n","Sigma.Gysq is:","\n","\n")
print(sigma.Gysq)

# Get revised values.

A1<-matrix(nrow=N,ncol=N)
        for (i in 1:N)
        A1[i, ]<-sigma.N[ ,i]/sigma.J
cat("\n","The revised A1 matrix is:","\n","\n")
print(A1)

g1<-(1/factor)*(sigma.Gy/sigma.J)
cat("\n","The revised g1 vector is:","\n","\n")
print(g1)

gamma1<-(1/factor)*((((sigma.Gysq)-(2*g1*sigma.Gy)
       +((g1^2)*sigma.J))/(sigma.J)))
cat("\n","The revised gamma1 vector is:","\n","\n")
print(gamma1)

gamma2<-(1/(dt)^(1/2))*(1/factor)*((((sigma.Gysq)-(2*g1*sigma.Gy)
       +((g1^2)*sigma.J))/(sigma.J)))
cat("\n","The revised gamma2 vector is:","\n","\n")
print(gamma2)

# Get g1.hat, gamma1.hat, and gamma2.hat.

g1.hat<-vector(length=K)
        for (k in 1:(length(g1.hat)))
        g1.hat[k]<-(g1%*%q.norm[ ,k])
cat("\n","g1.hat is:","\n","\n")
print(g1.hat[K])

gamma1.hat<-vector(length=K)
        for (k in 1:(length(gamma1.hat)))
        gamma1.hat[k]<-(gamma1%*%q.norm[ ,k])
cat("\n","gamma1.hat is:","\n","\n")
print(gamma1.hat[K])

gamma2.hat<-vector(length=K)
        for (k in 1:(length(gamma2.hat)))
        gamma2.hat[k]<-(gamma2%*%q.norm[ ,k])
cat("\n","gamma2.hat is:","\n","\n")
print(gamma2.hat[K])

rm(GAMMA.A,q.matrix,q0,q,N0,N.hat,J0,J,G0,Gy,G0.sq,Gysq)
rm(sigma.N,sigma.J,sigma.Gy,sigma.Gysq,gamma2,gamma2.hat,q.sum)
```

```
rm(N.check)
A<-A1
rm (A1)

# Test for normality.

vec<-qnorm(ppoints(K),mean=0,sd=1) # Calculate the normal scores.

test1.statistic<-vector(length=K)
        for (k in 1:(length(test1.statistic)))
        test1.statistic[k]<-((y[k]-g1.hat[k])/gamma1.hat[k])
cat("\n","The vector of test statistics, using g1 and gamma1, is:")
cat("\n","\n")
print(test1.statistic)

cat("\n","The mean and variance of this vector are:","\n","\n")
print(mean(test1.statistic))
print(var(test1.statistic))
rm(y,g1.hat,gamma1.hat)

cat("\n","The correlation between the values of the test statistic")
cat("\n","and its normal scores is:")
cat("\n","\n")
print(cor(vec,sort(test1.statistic)))
rm(vec,test1.statistic)

# Prediction.

pred1.price<-rep(0,K) # This is used to store the predicted prices.

pred1.mult<-rep(0,K) # This is used to store the factor that is
# multiplied by St to get E[S(t+1)].

for (k in 1:K) {

        pred1<-vector(length=N)
                for (i in 1:(length(pred1)))
                pred1[i]<-(exp(g1[i]/factor)*exp((1/2)*(1/factor^2)
                *(gamma1[i]^2))
                q.norm[ ,k][i])

        pred1.mult[k]<-sum(pred1)

        pred1.price[k]<-(prices[ ,b][k+1]*sum(pred1)) }

cat("\n","The vector of factors that is multiplied by St to get")
cat("\n","E[S(t+1)] is:","\n","\n")
print(pred1.mult)
cat("\n","The vector of predicted prices is:")
cat("\n","\n")
print(pred1.price)
rm(g,gamma,q.norm,pred1,pred1.mult)

cat("\n","The predicted price for the next period, S.hat(K+2)")
cat("\n","(i.e. the first price in the next batch of prices) is:")
cat("\n","\n")
```

```
print(pred1.price[K])
rm(K,pred1.price)

g<-g1
gamma<-gamma1
rm(g1,gamma1) }
```

Appendix B

Computer Program for Chapter 4

This appendix contains a copy of the **S** program written to implement the estimation procedure of Chapter 4.

```
# The values that must be entered are: the tau vector, B
# (the number of batches), K (the number of observations
# in each batch), TOLERANCE, factor, dt, m, and n.

# Also assign starting values to the g, gamma, and A matrices.

# The yield data are in a file called "yield" which has one column
# for each maturity value.

# Enter the data and parameter values.

mydata<-read.table("yield")

tau<-c(21/252,63/252,126/252,504/252,1260/252,1764/252,2520/252,
       7560/252)
cat("\n","The tau vector is:","\n","\n")
print(tau)

M<-(length(tau)) # Number of maturities.
cat("\n","The number, M, of maturities is:","\n","\n")
print(M)

B<-15 # Number of batches.
cat("\n","The number, B, of batches is:","\n","\n")
print(B)

K<-15 # Number of yields for each maturity within each batch.
cat("\n","The number, K, of yields for each maturity within each")
cat("\n"," batch is:")
cat("\n","\n")
print(K)

TOLERANCE<-(1.0e-10) # Elements in the GAMMA matrix are not allowed
# to fall below the level of TOLERANCE.
cat("\n","The specified TOLERANCE is:","\n","\n")
print(TOLERANCE)

factor<-5.0
cat("\n","The factor multiplying the y values is:","\n","\n")
print(factor)

# Enter the time step.

dt<-(5/252) # The time interval between yield quotes.
cat("\n","The time step is:","\n","\n")
print(dt)

# Initialize.

m<-2
n<-2
N<-m*n
#cat("\n","The value of m is:","\n","\n")
#print(m)
#cat("\n","The value of n is:","\n","\n")
#print(n)
```

```
cat("\n","The value of N is:","\n","\n")
print(N)

g<-matrix(nrow=N,ncol=M)
        for (i in 1:M)
        g[ ,i]<-c(0.9,0.1,0.8,0.2)
cat("\n","The g matrix is:","\n","\n")
print(g)

gamma<-matrix(nrow=N,ncol=M)
        for (m in 1:M)
        gamma[ ,m]<-c(0.25,0.25,0.25,0.25)
cat("\n","The gamma matrix is:","\n","\n")
print(gamma)

A<- matrix((1/N),nrow=N,ncol=N)
cat("\n","The A matrix is:","\n","\n")
print(A)

rm(m,n,p,q)

BP.matrix<-matrix(0,nrow=B,ncol=M)
# This is used to store the estimated prices calculated after
# processing each batch of yields.

y.matrix<-matrix(0,nrow=B,ncol=M)
# This is used to store estimated yields calculated after processing
# each batch of yields.

corr.matrix<-matrix(0,nrow=B,ncol=M)
# This is used to store the M correlations calculated for each of
# the B batches of yields.

# Run the batches of data.

for (b in 1:B) {

cat("\n","The batch number is:","\n","\n")
print(b)

y<-matrix(nrow=K,ncol=M)
        for (k in 1:K){
        for (m in 1:M)
        y[k,m]<-(factor*mydata[(K*(b-1)+k),m]/100) }
cat("\n","The y matrix is:","\n","\n")
print(y)

mean.y<-vector(length=M)
        for (m in 1:M)
        mean.y[m]<-mean(y[ ,m])
cat("\n","The means of the entries in each of the columns of the")
cat("\n","y matrix are:","\n","\n")
print(mean.y)

var.y<-vector(length=M)
        for (m in 1:M)
```

```
var.y[m]<-var(y[ ,m])
cat("\n","The variances of the entries in each of the columns of")
cat("\n","the y matrix are:","\n","\n")
print(var.y)

phi<-function(input) {
        (((2*pi*dt)^((-1)/2))*exp(((-1)*((input)^(2)))/(2*dt))) }

GAMMA1<-array(dim=c(K,N,M))
        for (k in 1:K) {
        for (n in 1:N) {
        for (m in 1:M) {
        GAMMA1[k,n,m]<-(phi((y[k,m]-g[n,m])/gamma[n,m])/
        (gamma[n,m]*phi(y[k,m])))
        if (GAMMA1[k,n,m]<TOLERANCE) GAMMA1[k,n,m]<-TOLERANCE )}}

GAMMA<-matrix(nrow=K,ncol=N)
        for (k in 1:K) {
        for (n in 1:N) {
        GAMMA[k,n]<-prod(GAMMA1[k,n, ])
        if (GAMMA[k,n]<TOLERANCE) GAMMA[k,n]<-TOLERANCE })
cat("\n","The GAMMA matrix is:","\n","\n")
print(GAMMA)

rm(GAMMA1)

e.matrix<-diag(N)

GAMMA.E<-array(dim=c(N,K,N))
        for (k in 1:K) {
        for (n in 1:N) {
        GAMMA.E[ ,k,n]<-GAMMA[k,n]*e.matrix[ ,n] })
rm(e.matrix,g,gamma,mean.y,var.y)

# Recurrences.

# Get q vectors.

q0<-rep((1/N),N)

Aq0<-(A%*%q0)

q.matrix<-matrix(0,nrow=N,ncol=(K+1))
q.matrix[ ,1]<-q0
# This is used to store the q vector after each iteration.
rm(q0)

q.norm<-matrix(0,nrow=N,ncol=K)
# This is used to store the normalized q vector after each iteration.
# This is our estimate of E[Xk/Yk], for k=1,2, ..., K.

for (k in 1:K) {

        q1<-matrix(nrow=N,ncol=N)
                for (n in 1:N) {
                q1[ ,n]<-(Aq0[n]*GAMMA.E[ ,k,n]) }
```

```
rm(Aq0)

        q<-vector(length=N)
                for ( n in 1:N)
                q[n]<-sum(q1[n, ])
        rm(q1)

        q.matrix[ ,(k+1)]<-q # Store q at the end of each iteration.

        q.norm[ ,k]<-(q.matrix[ ,k]/sum(q.matrix[ ,k]))

        Aq0<-(A%*%q) )

#       cat("\n","The q vector after K iterations is:","\n","\n")
#       print(q.matrix[ ,(K+1)])
        cat("\n","The normalized q vector, our estimate of E[Xk/Yk],")
        cat("\n","after K iterations is:","\n","\n")
        print(q.norm[ ,K])
        rm(Aq0,q)

# Get the N array.

AN0<-array(0,dim=c(N,N,N))

for (k in 1:K) {

        N1<-array(dim=c(N,N,N,N))
                for (r in 1:N) {
                for (s in 1:N) {
                for (n in 1:N) {
                N1[ ,r,s,n]<-AN0[ ,r,s][n]*GAMMA.E[ ,k,n] )))
        rm(AN0)

        N.hat<-array(dim=c(N,N,N))
                for (n in 1:N) {
                for (r in 1:N) {
                for (s in 1:N) {
                N.hat[n,r,s]<-if (n==s) (sum(N1[n,r,s, ])
                +(q.matrix[ ,k][r]*A[s,r]*GAMMA[k,s] ))
                else (sum(N1[n,r,s, ])) )))
        rm(N1)

        AN0<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (s in 1:N) {
                        AN0[ ,r,s]<-(A%*%N.hat[ ,r,s]) ))
        }

#       cat("\n","The N.hat array after K iterations is:","\n","\n")
#       print(N.hat)
        rm(GAMMA,AN0)

# Get J matrix.

AJ0<-matrix(0,nrow=N,ncol=N)
```

```
for (k in 1:K) {

        J1<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (n in 1:N) {
                J1[ ,r,n]<-AJ0[ ,r][n]*GAMMA.E[ ,k,n] })
        rm(AJ0)

        J2<-array(dim=c(N,N,N))
                for (r in 1:N) {
                for (n in 1:N) {
                J2[ ,r,n]<-A[n,r]*GAMMA.E[ ,k,n] })

        J<-matrix(nrow=N,ncol=N)
                for (n in 1:N) {
                for (r in 1:N) {
        J[n,r]<-(sum(J1[n,r, ])+(q.matrix[ ,k][r]*sum(J2[n,r, ]))) })
        rm(J1,J2)

        AJ0<-matrix(nrow=N,ncol=N)
                for (r in 1:N) {
                AJ0[ ,r]<-(A%*%J[ ,r]) )
        }

#       cat("\n","The J matrix after K iterations is:","\n","\n")
#       print(J)
        rm(AJ0)

# Get Gy array.

AG0<-array(0,dim=c(N,N,M))

for (k in 1:K) {

        Gy1<-array(dim=c(N,N,M,N))
                for (r in 1:N) {
                for (m in 1:M) {
                for (n in 1:N) {
                Gy1[ ,r,m,n]<-(AG0[ ,r,m][n]*GAMMA.E[ ,k,n]) }))
        rm(AG0)

        Gy2<-array(dim=c(N,N,M,N))
                for (r in 1:N) {
                for (m in 1:M) {
                for (n in 1:N) {
                Gy2[ ,r,m,n]<-A[n,r]*y[k,m]*GAMMA.E[ ,k,n] }))

        Gy<-array(dim=c(N,N,M))
                for (n in 1:N) {
                for (r in 1:N) {
                for (m in 1:M) {
                Gy[n,r,m]<-(sum(Gy1[n,r,m, ])+(q.matrix[ ,k][r]*
                sum(Gy2[n,r,m, ]))) }))
        rm(Gy1,Gy2)

        AG0<-array(dim=c(N,N,M))
```

```
for (r in 1:N) {
                for (m in 1:M) {
                AG0[ ,r,m]<-(A%*%Gy[ ,r,m]) }}
        }

#       cat("\n","The Gy array after K iterations is:","\n","\n")
#       print(Gy)
        rm(AG0)

# Get the Gysq array.

AG0.sq<-array(0,dim=c(N,N,M))

for (k in 1:K) {

        Gysq1<-array(dim=c(N,N,M,N))
                for (r in 1:N) {
                for (m in 1:M) {
                for (n in 1:N) {
                Gysq1[ ,r,m,n]<-AG0.sq[ ,r,m][n]*GAMMA.E[ ,k,n] }}}
        rm(AG0.sq)

        Gysq2<-array(dim=c(N,N,M,N))
                for (r in 1:N) {
                for (m in 1:M) {
                for (n in 1:N) {
                Gysq2[ ,r,m,n]<-A[n,r]*(y[k,m]^2)*GAMMA.E[ ,k,n] }}}

        Gysq<-array(dim=c(N,N,M))
                for (n in 1:N) {
                for (r in 1:N) {
                for (m in 1:M) {
                Gysq[n,r,m]<-(sum(Gysq1[n,r,m, ])+(q.matrix[ ,k][r]*
                sum(Gysq2[n,r,m, ])))}}}
        rm(Gysq1,Gysq2)

        AG0.sq<-array(dim=c(N,N,M))
                for (r in 1:N) {
                for (m in 1:M) {
                AG0.sq[ ,r,m]<-(A%*%Gysq[ ,r,m]) }}
        }

#       cat("\n","The Gysq array after K iterations is:","\n","\n")
#       print(Gysq)
        rm(AG0.sq,A,GAMMA.E,q.matrix)

# Get the totals.

sigma.N<-matrix(nrow=N,ncol=N)
        for (r in 1:N) {
        for (s in 1:N) {
        sigma.N[r,s]<-sum(N.hat[ ,r,s]) }}

sigma.J<-vector(length=N)
        for (r in 1:N)
        sigma.J[r]<-sum(J[ ,r])
```

```
sigma.Gy<-matrix(nrow=N,ncol=M)
        for (r in 1:N) {
        for (m in 1:M) {
        sigma.Gy[r,m]<-sum(Gy[ ,r,m]) })

sigma.Gysq<-matrix(nrow=N,ncol=M)
        for (r in 1:N) {
        for (m in 1:M) {
        sigma.Gysq[r,m]<-sum(Gysq[ ,r,m]) })

cat("\n","Sigma.N is:","\n","\n")
print(sigma.N)
cat("\n","Sigma.J is:","\n","\n")
print(sigma.J)
cat("\n","Sigma.Gy is:","\n","\n")
print(sigma.Gy)
cat("\n","Sigma.Gysq is:","\n","\n")
print(sigma.Gysq)
rm(N.hat,J,Gy,Gysq)

# Get the revised values.

A1<-matrix(nrow=N,ncol=N)
        for (n in 1:N) {
        A1[n, ]<-sigma.N[ ,n]/sigma.J }

g1<-matrix(nrow=N,ncol=M)
        for (m in 1:M) {
        g1[ ,m]<-((1/factor)*(sigma.Gy[ ,m]/sigma.J)) }

gamma1<-matrix(nrow=N,ncol=M)
        for (m in 1:M) {
        gamma1[ ,m]<-((1/factor)*(((sigma.Gysq[ ,m])-(2*g1[ ,m]*
        sigma.Gy[ ,m])+((g1[ ,m]^2)*sigma.J))/(sigma.J))) }

cat("\n","The revised A matrix is:","\n","\n")
print(A1)
cat("\n","The revised g1 matrix is:","\n","\n")
print(g1)
cat("\n","The revised gamma1 matrix is:","\n","\n")
print(gamma1)
rm(sigma.N,sigma.J,sigma.Gy,sigma.Gysq)

# Get g1.hat and gamma1.hat.

g1.hat<-matrix(nrow=K,ncol=M)
        for (k in 1:K) {
        for (m in 1:M) {
        g1.hat[k,m]<-(g1[ ,m]%*%q.norm[ ,k]) })

gamma1.hat<-matrix(nrow=K,ncol=M)
        for (k in 1:K) {
        for (m in 1:M) {
        gamma1.hat[k,m]<-(gamma1[ ,m]%*%q.norm[ ,k]) })
cat("\n","The g.hat vector is:","\n","\n")
print(g1.hat[K, ])
```

```
cat("\n","The gamma.hat vector is:","\n","\n")
print(gamma1.hat[K, ])

# Get the prices for the various maturities.

BP1<-matrix(nrow=N,ncol=M)
        for (m in 1:M) {
        BP1[ ,m]<-(payment*exp((-1)*g1[ ,m]*tau[m])) }

BP<-vector(length=M)
        for (m in 1:M)
        BP[m]<-(BP1[ ,m]%*%q.norm[ ,K])
#cat("\n","The vector of prices for the various maturities is:")
#cat("\n","\n")
#print(BP)

yld<-vector(length=M)
        for (m in 1:M)
        yld[m]<-(((-1)*log(BP[m]))/tau[m])
#cat("\n","The vector of estimated yields for the various")
#cat("\n","maturities is:","\n","\n")
#print(yld)

BP.matrix[b, ]<-BP
y.matrix[b, ]<-yld

rm(BP1,BP,q.norm,yld)

# Testing the model.

test.statistic<-matrix(nrow=K,ncol=M)
        for (m in 1:M) {
        test.statistic[ ,m]<-((y[ ,m]-g1.hat[ ,m])/gamma1.hat[ ,m]) }

mean.vector<-vector(length=M)
        for (m in 1:M) {
        mean.vector[m]<-(mean(test.statistic[ ,m])) }

var.vector<-vector(length=M)
        for (m in 1:M) {
        var.vector[m]<-(var(test.statistic[ ,m])) }

cat("\n","The matrix of test statistics is:","\n","\n")
#print(test.statistic)
cat("\n","The mean vector is:","\n","\n")
print(mean.vector)
cat("\n","The variance vector is:","\n","\n")
print(var.vector)

vec<-qnorm(ppoints(K),mean=0,sd=1) # Calculates the normal scores.

for (m in 1:M) {
        corr.matrix[b,m]<-(cor(vec,sort(test.statistic[ ,m]))))

rm(test.statistic,mean.vector,var.vector,g1.hat,gamma1.hat,y)
```

```
A<-A1
g<-g1
gamma<-gamma1
rm(A1,g1,gamma1)
}

cat("\n","The matrix of prices, where the bth row gives prices")
cat("\n","by maturity for the bth batch of data, is:")
cat("\n","\n")
print(BP.matrix)

cat("\n","The matrix of estimated yields, where the bth row gives")
cat("\n","estimated yields by maturity for the bth batch of")
cat("\n","data, is:","\n","\n")
print(y.matrix)
```